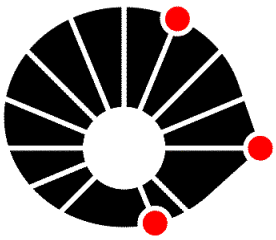
Universidade Estadual de Campinas
Faculdade de Engenharia Elétrica e Computação
Departamento de Engenharia de Computação e de Automação Industrial



Estudo por Simulação do Protocolo TCP de Alta Velocidade

Autor: Evandro de Souza
Orientador: Prof. Dr. Eleri Cardozo
Co-Orientadora: Dra Deborah Anne Agarwal

Banca Examinadora:

Prof. Dr. Eleri Cardozo
FEEC/UNICAMP

Prof. Dr. Michael Anthony Stanton
IC/UFF

Prof. Dr. Marco Aurélio Amaral Henriques
FEEC/UNICAMP

Prof. Dr. Lee Luan Ling
FEEC/UNICAMP

Dissertação apresentada ao Departamento de Engenharia de Computação e Automação Industrial da Faculdade de Engenharia Elétrica e de Computação da UNICAMP, como parte dos requisitos para obtenção do grau de Mestre em Engenharia Elétrica.

Campinas - SP - Brasil
Agosto, 2003

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

Souza, Evandro de
So89e Estudo por simulação do protocolo TCP de alta velocidade. / Evandro de Souza.–Campinas, SP: [s.n.],2003.

Orientadores: Eleri Cardozo; Deborah Anne Agarwal.
Dissertação (mestrado) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Redes de computadores. 2. Redes de longa distância (Redes de computação). I. Cardozo, Eleri. II. Agarwal, Deborah Anne. III. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. IV. Título.

Resumo

O atual mecanismo de controle de congestionamento usado no protocolo TCP tem dificuldade para atingir a completa utilização de enlaces de alta velocidade, particularmente em conexões de longa distância. Por exemplo, a taxa de descarte de pacotes necessária para preencher um canal de dados da ordem de Gigabits usando o atual protocolo TCP está abaixo da presente taxa de erros alcançável por fibra ótica. O TCP de Alta Velocidade foi proposto como uma modificação do mecanismo de controle de congestionamento do TCP visando atingir um desempenho razoável em enlaces de alta velocidade em longa distância. Nesta pesquisa são apresentados os resultados de simulação mostrando o desempenho do TCP de Alta Velocidade e o impacto do seu uso na atual implementação do TCP. Condições de rede incluindo diferentes níveis de congestionamento, diferentes níveis de taxa de perdas, diferentes graus de tráfego em rajadas e duas políticas distintas de gerenciamento de enfileiramento do roteador foram simuladas. O desempenho e a imparcialidade do TCP de Alta Velocidade foram comparados com o TCP existente e com soluções para a transferência volumosa de dados usando fluxos paralelos.

Palavras chave: TCP de Alta Velocidade, Controle de Congestionamento do TCP, Rede de Longa Distância, Transferência Volumosa de Dados, Redes de Computadores

Abstract

The current congestion control mechanism used in TCP has difficulty to reach full utilization on high speed links, particularly on wide-area connections. For example, the packet drop rate needed to fill a Gigabit pipe using the present TCP protocol is below the currently achievable fiber optic error rates. HighSpeed TCP was proposed as a modification of TCP's congestion control mechanism in order to achieve reasonable performance in high speed wide-area links. In this research, simulation results showing the performance of HighSpeed TCP and the impact of its use on the present implementation of TCP are presented. Network conditions including different degrees of congestion, different levels of loss rate, different degrees of bursty traffic and two distinct router queue management policies were simulated. The performance and fairness of HighSpeed TCP were compared to the existing TCP and solutions for bulk-data transfer using parallel streams.

Keywords: HighSpeed TCP, TCP Congestion Control, Wide Area Network, Bulk Data Transfer, Computer Networks

Agradecimentos

Este trabalho tem as suas bases na inspiração, encorajamento, ajuda e confiança de um número de pessoas, as quais eu gostaria de expressar a minha sincera gratidão.

Primeiro e sobre todos, eu agradeço a Deus que faz o impossível acontecer e mantém imutáveis as suas promessas.

Eu gostaria de expressar a meu agradecimento ao Dr. Eleri Cardozo pela imensa confiança na minha pessoa e no meu trabalho.

Eu gostaria dar uma agradecimento especial a Deborah Anne Agarwal por seu encorajamento, valiosas instruções e suporte em diversas áreas durante o meu período de pesquisa no Lawrence Berkeley National Laboratory.

A seguir, eu gostaria de extender os meus sinceros agradecimentos à Sally Floyd por suas inumeráveis e valiosas respostas, e inspiração para muitas das idéias apresentadas neste trabalho. Elas foram de grande ajuda para este estudo.

Também extendo este reconhecimento a Stanley Oliveria por seu esforço em ajudar-me a escrever minha tese. Ele foi uma grande fonte de encorajamento e apoio para mim. Expresso meus agradecimentos a Adailton Santos Silva pelas sugestões e críticas que enriqueceram este trabalho, bem como a Paulo César de Oliveira e Carlos Alberto Alves Meira por seu trabalho como conselheiros acadêmicos.

Além disso, eu gostaria de mencionar minha gratidão a minha família, em especial a minha amada esposa Nilcéia. Sem ela, nada disso seria possível. Eu extendo esta gratidão aos meus pais e sogros por seu amor e encorajamento.

Conteúdo

Lista de Figuras

Lista de Tabelas

Abreviaturas

ACK	Acknowledgment
AIMD	Additive Increase Multiplicative Decrease
AQM	Active Queue Management
BDP	Bandwidth Delay Product
BSD	Berkeley Software Distribution
ECN	Explicit Congestion Notification
FIFO	First In First Out
FTP	File Transfer Protocol
CWND	Congestion Window
DT	DropTail
DWDM	Dense Wavelength Division Multiplexing
HIPPI	High Performance Parallel Interface
HSTCP	HighSpeed TCP
HTTP	Hyper Text Transfer Protocol
MAX_SSTHRESH	Maximum Slow-Start Threshold
MSS	Maximum Segment Size
MTU	Maximum Transfer Unit
NS-2	Network Simulator Version 2
RED	Random Early Detection
REGTCP	Regular TCP
RFC	Request For Comments
RTO	Retransmit Timer
RTT	Round-Trip Time
SACK	Selective Acknowledgement
SSTHRESH	Slow-Start Threshold
SYN	Synchronization Packet
TCP	Transmission Control Protocol
XCP	Explicit Control Protocol
WWW	World Wide Web

Capítulo 1

Introdução

Uma das grandes questões na área de conectividade atualmente é a crescente demanda por uma maior capacidade de banda. Diversas tecnologias têm emergido e aumentado a capacidade dos canais de comunicação em várias vezes. Atualmente, vários meios estão disponíveis para conectar dois pontos em alta velocidade, tanto para conexões locais, como para ligações de longa distância.

O meio mais importante de conexão atualmente é a fibra ótica. Ela foi um marco crucial alcançado na evolução das redes de comunicação nos últimos 20 anos. Um recente desenvolvimento no campo das redes óticas produziu a técnica de *Dense Wavelength Division Multiplexing* (DWDM). Esta técnica realiza um processo de multiplexação de vários comprimentos de onda diferentes em uma única fibra ótica, abrindo um extraordinário meio para a transmissão de dados. Portanto, o ponto de gargalo está agora mudando do canal de comunicação para os equipamentos nos pontos finais de uma ligação, quando se deseja uma conexão de alto desempenho [27].

Com a chegada de aplicações que demandam grande largura de banda, tais como transferência de grande volume de dados (*bulk-data transfer*), transmissões multimídia e grades computacionais [1] para computação de alta performance, o desempenho da rede em conexões de longa distância tem se tornado um componente crítico da infra-estrutura [14].

Na Internet, o *Transmission Control Protocol* (TCP) tem sido amplamente usado como protocolo de transporte. Vários protocolos como o *Hyper Text Transfer Protocol* (HTTP) e o *File Transfer Protocol* (FTP) foram projetados para operar sobre o TCP. O TCP foi desenvolvido na década de 70 e tem sido constantemente modificado desde então para adaptar-se a novas apli-

[1] Um conjunto de computadores ligados através de ligações de longa distância que são capazes de realizar tarefas em conjunto.

cações, acomodar-se a características particulares dos novos meios de transmissão e também para melhorar sua capacidade de transferência de dados.

Observações recentes têm indicado que o mecanismo de controle de congestionamento do TCP tem dificuldades para atingir plena utilização dos canais óticos, particularmente em conexões de longa distância. As aplicações de rede raramente são capazes de utilizar completamente as novas redes de alta velocidade, bem como não utilizam plenamente toda a banda disponível [28]. A taxa de perda de pacotes necessária para uma transmissão utilizar completamente um canal de comunicação da ordem de gigabits por segundo, usando o atual protocolo TCP, está muito abaixo da atual taxa de erros das fibras óticas e, conseqüentemente, o controle de congestionamento torna-se não tão dinâmico a mudanças na condições de rede [23]. Sem o auxílio no diagnóstico e configuração das aplicações por parte de engenheiros especialistas em rede, a maioria dos usuários raramente irá atingir uma transmissão de 5 Mbps em uma transmissão de TCP, usando um único fluxo, apesar do fato de que a capacidade da infraestrutura de rede pode suportar taxas de transmissão de 100 Mbps ou mais [30].

Estas observações têm motivado várias pesquisas na área de redes de alta velocidade, objetivando melhorar o desempenho do TCP em situações onde existe um elevado produto banda atraso. Diversas propostas têm emergido na literatura tratando de alguns dos aspectos deste complexo problema [17, 39, 54, 35, 36].

Por outro lado, manter a utilização do recurso de banda de maneira eqüitativa entre múltiplas conexões homogêneas e heterogêneas em uma rede é um aspecto essencial [27]. Portanto, as novas soluções propostas não devem interferir com as soluções já existentes, ou somente interferir quando os protocolos existentes são incapazes de utilizar plenamente a capacidade disponível.

Este estudo pretende analisar experimentalmente o emprego de uma modificação proposta no mecanismo de controle de congestionamento do TCP para o uso em conexões com grandes janelas de congestionamento e baixa perda de pacotes, conhecida como TCP de Alta Velocidade (*HighSpeed TCP* ou HSTCP). Ele foi recentemente proposto e existem atualmente poucos estudos realizados sobre o seu uso [18].

A proposta geral deste trabalho é estudar a eficácia do TCP de Alta Velocidade em enlaces de alta velocidade e longa distância em regime estacionário, como um mecanismo para transferência de grande volume de dados, enquanto mantendo imparcialidade com outros tipos de TCP já em uso.

Na execução deste objetivo, as seguintes questões são abordadas:

- O comportamento do TCP de Alta Velocidade em situações nas quais o TCP Padrão possui baixo desempenho
- A possibilidade de utilizar o TCP de Alta Velocidade junto com o TCP Padrão e manter uma imparcialidade aceitável
- O efeito da política de enfileiramento do roteador no desempenho do TCP de Alta Velocidade e na imparcialidade entre o TCP de Alta Velocidade e o TCP Padrão
- A possibilidade do TCP de Alta Velocidade ser um substituto para outros tipos de transferência volumosa de dados

Estas questões são analisados em diferentes condições de rede. Estas condições incluem diferentes graus de congestionamento, diferentes níveis de taxa de perda de pacotes, vários níveis de tráfego em rajadas e duas políticas distintas de gerenciamento de enfileiramento do roteador. Espera-se que estas diferentes condições de rede apresentem uma ampla visão das virtudes e fraquezas do HSTCP.

A apresentação dos resultados desta pesquisa está organizada da seguinte forma. O Capítulo 2 apresenta uma breve história e os fundamentos do controle de congestionamento do TCP. Este Capítulo apresenta ainda os problemas atuais enfrentados pelo TCP para atingir um alto desempenho, bem como algumas soluções propostas para contornar estes obstáculos. O Capítulo 3 apresenta os fundamentos do HSTCP. O Capítulo 4 mostra a proposta para este trabalho, a abordagem seguida, bem como discute a metodologia usada nos experimentos e as métricas para a análise. Os resultados dos experimentos deste estudo estão descritos no Capítulo 5. O Capítulo 6 apresenta a discussão sobre os resultados obtidos e sua interpretação. O Capítulo 7 é dedicado para a conclusão e indicação de futuros trabalhos que podem seguir nesta linha de pesquisa.

Capítulo 2

Controle de Congestionamento do TCP

2.1 Visão Geral do TCP

O TCP é um protocolo projetado na década de 70. Muitos esforços de pesquisa, desenvolvimento e padronização têm sido extensivamente dedicados para a tecnologia TCP/IP. Ele é amplamente utilizado na atual Internet. Vários serviços da Internet como o *World Wide Web* (WWW) e FTP usam o TCP como protocolo da camada de transporte.

O TCP provê a entrega confiável de segmentos de dados através de um mecanismo de reconhecimento positivo. Cada segmento de dados transmitido contém um número de seqüência indicando a posição destes dados na transmissão. Ele é um protocolo *full-duplex*, significando que cada conexão TCP pode suportar um par de fluxos de bytes, sendo um em cada direção. O TCP também possui um mecanismo de multiplexação que permite que múltiplos aplicativos, em uma única máquina, realizem simultaneamente conexões com seus pares.

O TCP é um protocolo de janelas deslizantes. Um protocolo de janela deslizante permite que o transmissor envie um certo número de segmentos de dados antes de receber um Reconhecimento (*Acknowledgment* ou ACK). Quando um ACK é recebido pelo emissor, a janela *desliza* para permitir que mais um segmento de dados seja transmitido. Isto inclui um mecanismo de controle de fluxo para cada um dos fluxos de bytes, permitindo ao receptor limitar a quantidade de dados que o emissor pode transmitir de cada vez. Cada segmento TCP (segmento de dados e ACK) contém uma janela de anúncio (*advertised window*). O tamanho da janela de anúncio emitida pelo receptor é o limite superior para a janela deslizante do transmissor.

2.1.1 Colapso de Congestionamento

Nos primeiros anos de seu emprego, o TCP possuía apenas um mecanismo rudimentar de controle de congestionamento, que não era suficiente para prevenir o congestionamento em roteadores intermediários. Quando muitas conexões TCP transmitiam impropriamente a altas taxas, a rede sofria do chamado *Colapso de Congestionamento* (*Congestion Collapse*) [15]. O colapso de congestionamento é um estado no qual segmentos de dados são injetados na rede, porém muito pouco trabalho útil é conseguido, porque a maioria dos segmentos de dados e seus correspondentes ACKs são descartados por um dos roteadores intermediários da rede antes de atingir seus destinos. Isto faz com que o transmissor reenvie estes dados, agravando ainda mais o problema. Nagle [45] apresenta uma discussão detalhada sobre colapso de congestionamento. Os algoritmos de controle de congestionamento do TCP procuram prevenir o colapso de congestionamento através da detecção de congestionamento com a correspondente redução da taxa de transmissão.

Van Jacobson [31] demonstrou a importância do controle de congestionamento na Internet e propôs alguns dos algoritmos do TCP para evitar e controlar o congestionamento na rede. Este trabalho chamou a atenção de vários outros pesquisadores sobre a importância do controle de congestionamento do TCP. Lakshman [38] examinou o desempenho do TCP/IP em redes de longa distância. Paxson [48] investigou a dinâmica fim-a-fim da conexões na Internet incluindo os mecanismos do controle congestionamento do TCP e suas características. Como resultado destes esforços, várias RFC (*Request for Comments*) tratando do TCP foram publicadas para melhorar o seu desempenho [5, 32].

É notável que o TCP esteja ainda em uso atualmente, apesar do fato de ter sido desenvolvido a cerca de 25 anos atrás. O sucesso do TCP é devido principalmente à robustez do seu mecanismo de controle de congestionamento. Este mecanismo faz com que o TCP reduza sua taxa de transmissão quando algum congestionamento é observado através da perda de pacotes. O gerenciamento de congestionamento é imperativo como meio de permitir que a rede se recupere de congestionamentos e opere em um estado de baixa latência e alta transmissão. Na próxima seção serão apresentados os algoritmos mais importantes do controle de congestionamento do TCP.

2.1.2 Partida Lenta e Prevenção de Congestionamento

Proposto por Van Jacobson [31], os algoritmos de Partida Lenta (*Slow-Start*) e Prevenção de Congestionamento (*Congestion Avoidance*) permitem que o TCP aumente sua taxa de

transmissão sem sobrecarregar a rede. Eles utilizam uma variável chamada Janela de Congestionamento (*Congestion Window* ou CWND). A janela de congestionamento do TCP é do tamanho da janela deslizante usada pelo transmissor e não pode exceder a janela de anúncio do receptor. Portanto, o TCP não pode injetar na rede mais que CWND segmentos de dados sem notificação de recebimento.

O algoritmo de Partida Lenta é usado para aumentar a quantidade de dados sem confirmação de recebimento que o TCP injeta na rede, através do aumento gradual do tamanho da janela de congestionamento. A Partida Lenta é usada no começo de uma conexão TCP e, em certas circunstâncias, após a detecção de um congestionamento. O algoritmo começa inicializando a variável CWND em um segmento. Para cada ACK recebido, o TCP aumenta o valor de CWND em um segmento. Por exemplo, após a chegada do primeiro ACK, CWND é aumentada para dois segmentos, e o TCP é capaz de transmitir dois novos segmentos de dados. Este algoritmo habilita um crescimento exponencial no tamanho da janela de congestionamento. A Partida Lenta continua até que, ou o tamanho de CWND alcança o valor do Limiar da Partida Lenta (*Slow-Start Threshold* ou SSTHRESH) ou quando a perda de um segmento de dados é detectada, quando então é encerrada.

O valor de CWND é inicializado no valor do Tamanho Máximo de Segmento (*Maximum Segment Size* ou MSS). Este valor de MSS é baseado no MSS do receptor obtido durante a estabelecimento da conexão TCP, na Unidade Máxima de Transferência (*Maximum Transfer Unit* ou MTU) do caminho da conexão, no MTU da interface do transmissor; ou na ausência de outra informação, 536 bytes.

Se o receptor estiver enviando um ACK para cada pacote, o efeito deste algoritmo será de dobrar a taxa de envio de dados pelo transmissor a cada intervalo de Tempo de Percurso (*Round-Trip Time* ou RTT). Obviamente isto não pode ser sustentado indefinidamente. Ou o valor de CWND irá exceder a janela de anúncio do receptor, ou a janela do transmissor, ou a capacidade da rede será ultrapassada, causando perda de pacotes.

O outro limite para o aumento de CWND durante a Partida Lenta é mantido pela variável SSTHRESH. Se o valor de CWND aumentar, passando do valor de SSTHRESH, o controle de fluxo do TCP é mudado do algoritmo de Partida Lenta para o algoritmo de Prevenção de Congestionamento (*Congestion Avoidance*). Inicialmente o valor de SSTHRESH é colocado para o valor correspondente ao tamanho máximo da janela do receptor. Entretanto, quando um congestionamento é percebido, SSTHRESH é reduzido para a metade da janela atual, fornecendo ao TCP uma memória do ponto aonde ele poderá antecipar o congestionamento da rede no futuro.

O aumento da CWND durante a fase de Partida Lenta é interrompido quando a janela

de congestionamento exceder a janela de anúncio do receptor, quando a taxa de transmissão exceder o valor de congestionamento memorizado em SSTHRESH ou quando estiver além da capacidade da rede.

Quando a taxa de envio é maior que o nível que pode ser sustentado pela rede, pacotes de dados podem ser descartados por transbordamento de *buffer*. O TCP pode detectar a perda de pacotes de dois modos. Primeiro, se um único pacote é perdido dentro de uma seqüência de pacotes, a entrega bem sucedida dos pacotes posteriores ao pacote perdido fará que o receptor gere um ACK duplicado para cada entrega bem sucedida de um pacote. A chegada destes ACKs duplicados é um sinal de que houve perda de um pacote. Segundo, se um pacote é perdido no final de uma seqüencia de pacotes enviados, não há pacotes consecutivos para a geração de ACKs duplicados. Neste caso, não há a geração do correspondente ACK deste pacote. Em conseqüência, o Temporizador de Retransmissão (*Restransmit Timer* ou RTO) do transmissor irá expirar e o transmissor assumirá a perda do pacote.

Em um protocolo baseado em ACKs, o transmissor é responsável pela detecção de perdas de pacote. Pacotes perdidos são revelados pelas falhas na ordem dos números de seqüencia com confirmação de recebimento, devido a ausência de ACKs.

Prevenção de Congestionamento é a fase que segue a Partida Lenta. Nesta fase o valor da CWND é maior ou igual a SSTHRESH. Este algoritmo incrementa a CWND a uma taxa mais lenta do que durante a Partida Lenta. Para cada segmento com confirmação de recebimento durante a Prevenção de Congestionamento, a janela de congestionamento é incrementada em 1/CWND (a não ser que este incremento faça com que o valor de CWND fique maior que janela de anúncio do receptor). Isto adiciona aproximadamente um segmento ao valor de CWND a cada RTT. O algoritmo de Prevenção de Congestionamento provê um acréscimo na janela deslizante do TCP. Este mecanismo é usado para verificar se existe capacidade adicional na rede, de maneira conservadora.

A janela de congestionamento continua a aumentar desta maneira até que uma perda de pacote ocorra. Quando a perda de pacote acontece, os ACKs duplicados resultantes irão fazer com que transmissor corte pela metade a taxa de envio e continue o crescimento da janela de congestionamento a partir deste novo ponto.

A característica geral do algoritmo do TCP é de uma busca inicial rápida da capacidade da rede para estabelecer aproximadamente os limites de máxima eficiência, seguida de um comportamento adaptativo cíclico, que reage rapidamente a um congestionamento e incrementa lentamente a taxa de envio perto da área de máxima eficiência de transmissão. A perda de pacote, quando sinalizada pelo acionamento do RTO, faz com que o transmissor recomece a fase

de Partida Lenta, após um intervalo de timeout. Este comportamento geral é observado na Figura 2.1.

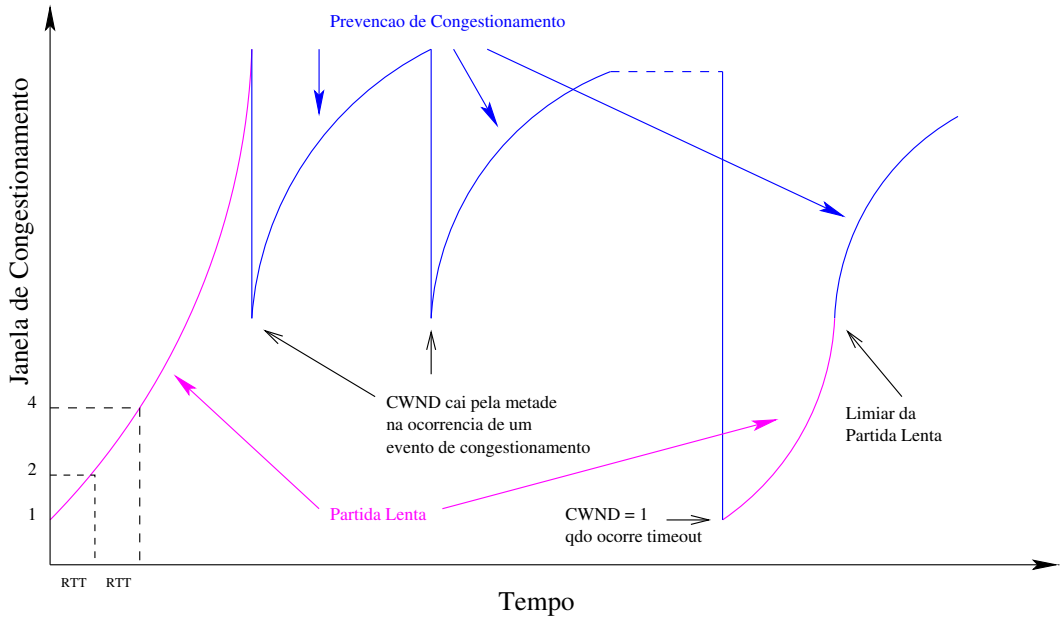


Figura 2.1: Controle de Congestionamento do TCP

Os algoritmos de controle de congestionamento também são conhecidos por Incremento Aditivo Decremento Multiplicativo (*Additive Increase Multiplicative Decrease* ou AIMD) e são a base do Controle de Congestionamento do TCP. Ele incrementa a janela de congestionamento em um pacote por janela de dados com confirmação de recebimento e corta pela metade a janela a cada janela de dados contendo um pacote perdido. De maneira simples, o controle de congestionamento do TCP pode ser expresso pelas seguintes equações:

Prevenção de Congestionamento

$$\textbf{ACK} \quad : \quad CWND \leftarrow CWND + \frac{a}{CWND} \tag{2.1}$$

$$\textbf{DROP} \quad : \quad CWND \leftarrow CWND - b \times CWND \tag{2.2}$$

Partida Lenta

$$\textbf{ACK} \quad : \quad CWND \leftarrow CWND + c \tag{2.3}$$

Os termos CWND, a e c são definidos em unidades de MSS. Os valores canônicos para a, b e c são: $a = 1$, $b = 0.5$ e $c = 1$.

Algumas melhorias importantes foram incorporadas ao controle de congestionamento do TCP, desde o trabalho inicial de Van Jacobson em 1988, que afetam o comportamento do TCP em conexões de alta velocidade. *Delay Acknowledment* permite que o receptor envie um ACK cumulativo de volta para o transmissor após ter recebido um número pré-definido de segmentos, ao invés de ter de reconhecer o recebimento de cada segmento. *Fast Retransmit* define que quando são recebidos três ACKs duplicados consecutivamente, o transmissor supõe que o pacote correspondente foi perdido e o retransmite, sem esperar que o RTO expire. *Fast Recovery* é usado para evitar a entrada na fase de Partida Lenta após cada perda de pacote. *Window Scale Option* aborda a questão do tamanho máximo de janela em situações que o caminho de rede possui alto produto banda atraso. *Timestamp Option* permite que o transmissor calcule mais precisamente o RTT para cada ACK recebido. Estas melhorias estão explicadas com maior detalhe em [56].

Uma melhoria recente e importante foi a introdução da opção de Reconhecimento Seletivo (*Selective Acknowledgment* ou SACK). Esta opção altera o comportamento da confirmação de recebimento do TCP. Esta opção de SACK é oferecida ao receptor durante a inicialização do TCP como uma opção do pacote de Sincronização (*Synchronization* ou SYN). O comportamento padrão da confirmação de recebimento do TCP é de reconhecer o maior número de seqüencia dos bytes ordenados. Este comportamento padrão está sujeito a causar retransmissões desnecessárias de dados, o que pode exacerbar uma condição de congestionamento que pode ter sido a causa da perda do pacote originalmente. A opção de SACK permite ao receptor modificar o campo de confirmação de recebimento para descrever blocos descontínuos de dados recebidos, de forma que o transmissor pode reenviar somente o que esteja faltando para o receptor [42].

Os algoritmos de controle de congestionamento para o SACK são uma extensão conservadora do controle de congestionamento do TCP Reno [13]. Eles utilizam os mesmos algoritmos para o aumento e diminuição da janela de congestionamento e atuam melhor quando um grande número de pacotes são descartados de uma janela de dados.

Os resultados do desempenho do TCP SACK baseado em simulações para caminhos com baixo e alto atraso foram documentados em [13]. Estes resultados sugeriram que o TCP SACK pode melhorar significativamente o desempenho na rede quando comparado com implementações anteriores do TCP.

2.2 Desempenho do TCP em Ligações de Alta Velocidade

Esta seção apresentará os principais problemas enfrentados pelo TCP para alcançar um alto desempenho em transmissões volumosas de dados em ligações de alta velocidade.

A introdução de tecnologias de redes de alta velocidade tem causado uma mudança dramática no desempenho das aplicações baseadas em TCP. É bem sabido que o desempenho do TCP depende da banda da rede, do tempo de percurso e da perda de pacotes. Quando a velocidade de uma ligação aumenta, a possibilidade de alcançar o produto banda atraso diminui. Este problema tende a aparecer mais em ligações transcontinentais de alta velocidade, muito embora esteja também aparecendo atualmente em redes locais [7].

Antes de expor este problemas, é necessário introduzir formalmente o conceito de produto banda atraso. A capacidade de uma conexão é normalmente medida em termos de Produto Banda Atraso (*Bandwidth Delay Product* ou BDP) [56]. Isto é expresso na forma:

$$\mathbf{Capacidade}_{(\mathbf{bits})} = \mathbf{banda}_{(\mathbf{bits/seg})} \times \mathbf{atraso}_{(\mathbf{seg})}$$

O atraso em uma rede é a latência de ida e volta necessária para a informação propagar do nó transmissor para o receptor. Banda é o número de bits que podem ser transmitidos em um certo período de tempo. BDP é o produto das métricas de desempenho acima, isto é, o número de bits (ou bytes) que uma rede pode suportar. É possível imaginar a rede como uma tubulação, na qual seu comprimento representa o tempo de latência de ida e volta da rede e a espessura representa a banda da conexão. Desta forma, BDP representa o volume da tubulação. Nas camadas de transporte e enlace, o BDP representa a quantidade máxima de dados sem confirmação de recebimento pendentes na rede em qualquer momento, para manter a conexão cheia. O desempenho do TCP não depende da taxa de transferência em si mesma, mas do produto da taxa de transferência pelo atraso de ida e volta. O BDP, portanto, mede a quantidade de dados que pode encher a *tubulação*. Quanto maior for o BDP, mais tempo levará para o TCP utilizar a capacidade disponível.

Partida Lenta

Para conexões TCP que são capazes de usar grandes janelas de congestionamento (de milhares de pacotes), o atual algoritmo de Partida Lenta pode resultar no aumento da janela de congestionamento em milhares de segmentos em um único RTT. Tal aumento pode facilmente resultar em milhares de pacotes sendo descartados em um único RTT. Isto é contra-producente para um fluxo TCP, bem como penoso para o restante do tráfego que compartilha uma conexão congestionada.

Este descarte de um grande número de pacotes pode resultar em timeouts de retransmissão desnecessários para uma conexão TCP. A conexão TCP pode entrar na fase de prevenção de congestionamento com uma janela de congestionamento muito pequena, bem como levar um grande número de RTTs para recuperar sua antiga janela de congestionamento [17].

Tamanho de Quadro

Atualmente, o tamanho típico do MSS do TCP é de 1448 bytes, devido ao valor de 1500 bytes da MTU de uma interface Ethernet. Este tamanho é útil para ser usado em múltiplas velocidades, bem como em ambientes de hubs e switches. Todavia, ele cria dificuldades para aplicações que necessitam enviar uma grande quantidade de dados, tais como transmissões volumosas de dados via FTP, porque estas aplicações trabalham melhor com quadros maiores. Um MSS maior melhora a velocidade de recuperação do TCP e reduz a taxa de interrupção da CPU por pacote. Outras mídias já suportam MTU maiores: FDDI tem um MTU de 4392 bytes, Gigabit Ethernet Jumbo Frame tem uma MTU de 9000 bytes, IP-over-ATM usa uma MTU de 9180 bytes e HiPPI usa uma MTU de 65535 bytes. Entretanto, todas estas mídias estão sendo deixadas de lado em favor do Ethernet de maior velocidade [7, 11].

Buffers TCP

Os nós transmissor e receptor requerem um espaço de *buffer* para lidar com os pacotes de chegada e partida em uma conexão. Este espaço deve ser de pelo menos a quantidade de dados sem confirmação de recebimento que o TCP precisa lidar de forma a manter o canal de comunicação cheio. Problemas de desempenho do TCP surgem quando o espaço para *buffer* não é adequado para acomodar o produto banda atraso. Se os *buffers* são muito pequenos, a janela de congestionamento do TCP nunca abrirá completamente a ponto de encher a conexão [39].

A janela de anúncio do receptor também precisa ser grande o suficiente. Ela limita o quanto o transmissor pode enviar, porque o transmissor não pode enviar mais dados do que a janela de anúncio permite. Todavia, a janela de congestionamento máxima é proporcional à quantidade de espaço de *buffer* que o núcleo do sistema operacional aloca para cada *socket*. Por exemplo, se o RTT é de 50ms e a banda desta conexão for 100 Mbits/seg, o *buffer* TCP deverá ser de 625.000 bytes. Como a capacidade de transmissão da rede aumentou nos últimos anos, os sistemas operacionais tem gradualmente modificado o tamanho de *buffer* padrão, normalmente congelado em valores entre 8 KBytes a 64 KBytes. Entretanto, estes valores ainda são muito pequenos para as redes de alta velocidade atuais [11], e impedem que o TCP faça uso de toda a banda disponível.

Algoritmo de Prevenção de Congestionamento

Atualmente, as implementações do TCP somente podem alcançar grandes janelas de congestionamento, quando houver uma taxa de perda de pacotes muito baixa. Perdas randômicas acarretam uma significativa deterioração da capacidade de transmissão quando o produto da probabilidade de perda pelo quadrado do produto banda atraso for maior que um [38]. Por exemplo, para uma transmissão com TCP padrão de pacotes com 1500 bytes e tempo de percurso de 100 ms, atingir uma taxa de transmissão de 10 Gbps seria necessária uma janela de congestionamento média de 83.333 segmentos e uma taxa de descarte de pacotes de, aproximadamente, um evento de congestionamento a cada 5.000.000.000 pacotes, ou equivalentemente a, no máximo, um evento de congestionamento a cada 1h:40m, conforme demonstrado em [18]. Isto está muito além do que é possível hoje em dia com a atual tecnologia de fibras óticas e de roteadores.

Buffers de Rede

O TCP tem, por natureza, característica de rajada. Esta característica de rajada do TCP pode resultar num fraco desempenho devido a uma limitada bufferização da rede. Grandes rajadas de dados colocadas na rede em curtos intervalos tendem a criar longas filas nos roteadores intermediários. Na maioria dos casos práticos, o tamanho máximo da janela, que reflete o maior tamanho possível de uma rajada de dados, é muito maior que a capacidade de enfileiramento de qualquer roteador intermediário. Uma vez que os transmissores TCP sobrecarreguem as filas dos roteadores, eles começarão a descartar pacotes. O TCP irá assumir estes descartes de pacotes, devido a este gargalo nas filas, como devido a congestionamento na rede. Isto pode resultar num fraco desempenho do TCP, com baixa taxa de transmissão e compartilhamento de banda despro-

porcional. Por outro lado, grandes filas no roteador podem introduzir atrasos adicionais nos fluxos TCP, aumentando seus RTT [55].

Nossa pesquisa está primariamente focada em problemas relacionados com o algoritmo de Prevenção de Congestionamento. Entretanto, os problemas mencionados anteriormente estão estreitamente relacionados, e podem ter tido certo impacto durante o desenvolvimento deste estudo.

2.3 Soluções Propostas para o Desempenho do TCP em Ligações de Alta Velocidade

Após ter apresentado os problemas enfrentados pelo TCP em ligações de alta velocidade, iremos rever a pesquisa que tem sido feita para superá-los. O resultado destas pesquisas estão apresentados nos ítens seguintes.

Partida Lenta Limitada

Partida Lenta Limitada (*Limited Slow-Start*) [17] é uma modificação no algoritmo de Partida Lenta do TCP para conexões TCP com grandes janelas de congestionamento. O *Limited Slow-Start* introduz um parâmetro chamado Limiar Máximo para a Partida Lenta (*Maximum Slow-Start Threshold* ou MAX_SSTHRESH). O algoritmo de Partida Lenta somente é modificado para valores de janela de congestionamento maiores de MAX_SSTHRESH. O algoritmo pode ser expresso da seguinte forma:

Para cada ACK chegando na Partida Lenta:

$$
\begin{aligned}
&if(CWND \leq MAX\_SSTHRESH) \\
&\qquad CWND = CWND + MSS; \\
&else \\
&\qquad K = int(CWND/(0.5 * MAX\_SSTHRESH)); \\
&\qquad CWND = CWND + int(MSS/K);
\end{aligned}
$$

Portanto, durante o *Limited Slow-Start*, a janela é acrescida de 1/K MSS para cada ACK recebido, na qual K = int(CWND/(0.5 * MAX_SSTHRESH), ao invés de 1 MSS como era o caso da Partida Lenta padrão [1]. O crescimento ainda será exponencial, porém mais lento. Quando SSTHRESH $<$ CWND, a Partida Lenta termina e o transmissor

passa para a fase de Prevenção de Congestionamento.

Tamanho de Quadro

Extensão do MTU

A referência [43] mostra que a taxa de transferência do TCP possui um limite superior baseado nos seguintes parâmetros:

$$Taxa\_de\_Transferencia \approx \frac{1.2 \times MSS}{RTT \times \sqrt{taxa\_perda\_de\_pacotes}}$$

Portanto, a taxa de transferência máxima do TCP é diretamente proporcional ao MSS, o qual é MTU menos os cabeçalhos dos pacotes TCP. Se todos os demais elementos forem iguais, é possível dobrar a taxa de transferência dobrando o tamanho do pacote. A taxa de perda de pacotes pode também aumentar com o tamanho de MSS, mas o faz a taxas sub-lineares, e em qualquer caso, possui o efeito do inverso do quadrado na taxa de transferência; ou seja, o tamanho MSS ainda irá dominar sobre a taxa de transferência [12]. Existem propostas para que haja um tamanho de quadro ainda maior (chamado *Jumbo*), especialmente para Gigabit Ethernet de 9000 bytes, ao invés dos atuais 1500 bytes, que são o atual padrão para tamanho de quadro Ethernet.

Fluxos Paralelos

Para melhorar o desempenho fim-a-fim, fluxos TCP paralelos podem ser usados [39]. Esta técnica é implementada dividindo-se os dados a serem transferidos em N porções e transferindo-se cada porção em conexões TCP separadas. Quando competindo com conexões em um enlace congestionado, cada um dos fluxos paralelos será menos passível de ser selecionado para ter seus pacotes descartados e, portanto, a quantidade potencial de banda agregada que precisará passar prematuramente para Prevenção de Congestionamento e Partida Lenta é reduzida. Uma aplicação abrindo N múltiplas conexões TCP está essencialmente criando um grande *MSS virtual* na conexão agregada, a qual é N vezes o MSS de uma única conexão [25].

Experimentos tem mostrado que fluxos paralelos podem melhorar dramaticamante a taxa de transferência [39, 58], mas esta abordagem tem sido considerada agressiva e não provê meios para uma divisão eqüitativa da banda de rede disponível para as aplicações [19].

Gerenciamento de *Buffer*

Sintonia Automática do Buffer do TCP

A sintonia automática de *buffer* TCP foi inicialmente proposta em [54]. Ela ajusta dinamicamente os *buffers* do *socket* do TCP para atingir taxas de transferência máxima em cada conexão TCP sem configuração manual. Ela é baseada nas condições de rede e na disponibilidade de memória do sistema. Como o produto banda atraso na Internet pode expandir-se em 4 ordens de magnitude, não é possível ter um único tamanho de *buffer* para todas as conexões em uma única máquina. Se os *buffers* forem sintonizados, é possível evitar o desperdício de memória do kernel, no caso do *buffer* ser muito grande. Por outro lado, também é possível evitar baixas taxas de transferência, quando o *buffer* de envio for pequeno. Neste esquema de sintonia, a janela do controle de fluxo do transmissor é ajustada.

Dynamic Right Sizing

Esta proposta é outra técnica de gerenciamento de *buffer* [14]. O receptor faz a estimativa da banda através da quantidade de dados recebidos a cada tempo de percurso. Esta estimativa é utilizada para mudar dinamicamente a janela de anúncio do receptor e também alocar mais imparcialmente os *buffers* para as conexões baseado na necessidade de cada conexão por *buffer*. O crescimento da janela de congestionamento do transmissor estará limitado pela quantidade de banda disponível.

Sintonia de Buffer do Linux

O *kernel* 2.4 do Linux inclui um algoritmo para a sintonia de *buffer*. Para aplicações que não indicarem explicitamente o tamanho dos *buffers* de recepção e transmissão, o *kernel* tentará fazer crescer o tamanho da janela para adequar-se com a banda disponível (até o limite da janela padrão do receptor). Se houver uma alta demanda por memória do *kernel* ou de rede, o tamanho do *buffer* pode ser limitado ou mesmo encolher.

Este processo é controlado pelas novas variáveis do *kernel* net.ipv4.tcp_rmem/wmem e pela quantidade de memória de *kernel* disponível [29].

Prevenção de Congestionamento

XCP

O Protocolo Explícito de Controle (*eXplicit Control Protocol* ou XCP) [35], generaliza a proposta da Notificação Explícita de Congestionamento (*Explicit Congestion Notification* ou ECN) [53]. Ao invés de um bit para indicação de congestionamento usado pelo ECN, roteadores habilitados para XCP informam aos transmissores sobre o grau de congestionamento no enlace gargalo. Cada pacote XCP carrega um cabeçalho de congestionamento, o qual é usado para para comunicar o estado do fluxo para os roteadores e realimentar informação dos roteadores para os receptores. Um campo informa a janela de congestionamento corrente do transmissor e outro comunica a estimativa presente de RTT do transmissor. Esta informação é preenchida pelo transmissor e não é modificada em trânsito. O terceiro campo é inicializado pelo transmissor e recebe realimentações dos roteadores ao longo do caminho para controlar diretamente as janelas de congestionamento das fontes. Do mesmo modo que o TCP, o XCP é um protocolo de controle de congestionamento baseado em janelas, projetado para tráfego de melhor esforço.

O XCP desassocia o controle de utilização do controle de imparcialidade. Para controlar a utilização, este protocolo ajusta sua agressividade de acordo com a banda livre na rede e com o valor de atraso. Para controlar a imparcialidade, o protocolo recupera banda dos fluxos cuja taxa de transferência esteja acima da porção justa e realoca-a para outros fluxos. A proposta do XCP reivindica ser estável e eficiente independente da capacidade do enlace, tempo de percurso e do número de fontes.

FAST TCP

O controle de congestionamento consiste de dois componentes, um algoritmo do transmissor, implementado no TCP, que adapta a taxa de envio à informação de congestionamento no seu caminho, e um algoritmo de enlace, implementado nos roteadores, que atualiza e realimenta a medida de congestionamento de volta para os transmissores que atravessam o enlace. Tem sido comprovado que os atuais algoritmos podem se tornar instáveis quando o atraso aumenta e mesmo quando a capacidade de rede

aumenta. O *FAST TCP* [33] propõe que, para manter a estabilidade, os transmissores devem diminuir sua resposta de acordo com o seu tempo de percurso individual e os enlaces devem diminuir suas respostas de acordo com a sua capacidade individual. Os autores reivindicam que estas duas ações, combinadas com a proposta de um modelo dual para o TCP Vegas [41], mantêm a estabilidade linear sem ter necessidade de trocar o atual algoritmo de enlace. Os autores da proposta implementaram um *kernel* FAST TCP com estes avanços e algumas características: ele usa tanto o atraso de enfileiramento quanto a perda de pacotes como sinais de congestionamento; ele trabalha com enormes perdas de pacotes; ele reduz a característica de rajada e as perdas excessivas de pacotes usando *pacing* no transmissor; e ele converge rapidamente para a vizinhança do valor de equilíbrio e então avança suavemente para o alvo.

Buffer de Rede

Paced TCP

O *Paced TCP* é uma modificação no TCP que procura resolver o problema do gargalo de enfileiramento que acontece quando existe um descasamento entre redes de alta capacidade e o armazenamento disponível individualmente nas filas dos roteadores de rede [36]. Um transmissor usando o *Paced TCP* libera os pacotes em múltiplas e pequenas rajadas durante um tempo de percurso, ao invés de liberar uma única grande rajada de pacotes, como o TCP padrão faz. Esta abordagem permite o transmissor aumentar sua taxa de envio até a janela máxima sem encontrar um gargalo de enfileiramento durante a Partida Lenta.

Capítulo 3

Fundamentos do TCP de Alta Velocidade

3.1 Descrição

O TCP de Alta Velocidade para Grandes Janelas de Congestionamento foi introduzido por Sally Floyd et al. [18] como uma modificação no mecanismo do controle de congestionamento do TCP para ser usado em conexões TCP com grandes janelas de congestionamento. Ele supera a dificuldade do TCP Padrão em atingir grandes janelas de congestionamento em ambientes com taxas de perda de pacotes muito baixas. O TCP de Alta Velocidade propõe uma pequena modificação nos parâmetros de incremento e decremento do TCP.

Num ambiente de estado estacionário, com uma baixa taxa de perda de pacotes p, a janela de congestionamento média do TCP é aproximadamente $1.2/sqrt(p)$ segmentos [19]. Isto coloca uma séria restrição na janela de congestionamento que pode ser atingida pelo TCP num ambiente real. Um exemplo do resultado desta limitação está descrito na seção 2.2, quando o algoritmo de Prevenção de Congestionamento é tratado. Se o tempo de percurso for maior, o tempo entre uma perda de pacote e a próxima seria ainda maior.

O TCP de Alta Velocidade não modifica o comportamento do TCP em ambiente com taxa de perda de pacote entre 1% e 5% e, portanto, não traz novas ameaças de

colapso de congestionamento. Ele foi projetado para ter uma resposta diferente em ambientes com taxa de perda de pacote muito baixa e ter a mesma resposta do TCP Padrão em ambientes com taxa de perda de pacotes de no mínimo 10^{-3}. Em ambientes com taxa de perda de pacote baixa (tipicamente inferiores a 10^{-3}) é possível ignorar elementos mais complexos da função resposta que são requeridos para modelar o desempenho do TCP em ambientes mais congestionados e com *timeouts* de retransmissão.

O TCP Padrão aumenta a sua janela de congestionamento em um segmento por janela de dados reconhecida e diminui a janela de congestionamento pela metade em cada janela de dados contendo um pacote descartado, seguindo o algoritmo clássico do AIMD. Na fase de Prevenção de Congestionamento, seu comportamento é expresso pelas equações 2.1 e 2.2 da página 9.

O número de tempos de percurso entre os eventos de congestionamento requeridos para um fluxo do TCP Padrão atingir uma taxa de transferência média alta aumenta diretamente com a banda disponível.

O número de tempos de percurso entre eventos de congestionamento para um fluxo TCP com tamanho de pacote D bytes, em função da taxa de transmissão média da conexão B em bits/seg, pode ser calculado através da janela de congestionamento média w de $BR/(8D)$, sendo R o tempo de percurso em segundos. Em estado estacionário, a janela de congestionamento média do TCP é aproximadamente $1.2/sqrt(p)$. Isto equivale a um evento de perda a cada $1/p$ pacotes, ou a cada $1/(pw) = w/1.5$. Substituíndo w, isto representa um evento de congestionamento a cada $(BR)/(12D)$ tempos de percurso.

Para pacotes de 1500 bytes e tempo de percurso (RTT) de 0.1 segundos, os números para a janela de congestionamento e taxa de perda de pacotes são apresentados na Tabela 3.1. Esta tabela também fornece a janela de congestionamento média w e a taxa de eventos de congestionamento em estado estacionário p.

Taxa de Transferência (Mbps)	RTTs Entre Perdas (segs)	Janela de Cong. (pcts)	Taxa de Perda
1	5.5	8.3	0.02
10	55.5	83.3	0.0002
100	555.5	833.3	0.000002
1000	5555.5	8333.3	0.00000002
10000	55555.5	83333.3	0.0000000002

Tabela 3.1: RTTs entre Perdas de Pacotes para o TCP Padrão

Pode-se perceber que para se atingir uma taxa de transferência elevada, é necessária uma taxa de perda de pacote muito pequena e um grande número de tempos de percurso entre as perdas, algo difícil de ser atingido em redes reais atualmente.

3.2 Função Resposta Modificada

A função resposta do TCP Padrão, $w = 1.2/sqrt(p)$, fornece uma janela de congestionamento média w para o TCP como sendo função da taxa de perda de pacotes em estado estacionário p. Esta função resposta do TCP Padrão é uma conseqüência direta dos mecanismos do AIMD.

O TCP de Alta Velocidade faz uso de uma função resposta diferente e fornece uma nova relação entre a janela de congestionamento média w e a taxa de perda de pacotes em estado estacionário p. Por simplicidade, esta nova função resposta do TCP de Alta Velocidade mantém a propriedade de que a função resposta fornece uma linha reta em escala log-log (assim com o faz a função resposta do TCP Padrão, para taxas de perda de pacote inferiores a 1%). Ambas funções resposta estão presentes na Figura 3.1.

A função resposta do TCP de Alta Velocidade é especificada usando-se três parâmetros: Low_Window, High_Window e High_P. Low_Window é usado para estabelecer um ponto de transição e assegurar compatibilidade. O TCP de Alta Velocidade usa a mesma função resposta que o TCP Padrão quando a janela de congestionamento atual é no máximo Low_Window e usa a função resposta do TCP de Alta Velocidade quando a janela de congestionamento atual é maior do que Low_Window. High_Window e High_P são usados para especificar o ponto superior da função resposta do TCP de Alta Velocidade. Ele é configurado a uma taxa de eventos de congestionamento específica High_P, necessária para a função resposta do TCP de Alta Velocidade atingir a janela de congestionamento média de High_Window.

A função resposta do TCP de Alta Velocidade pode ser traduzida nos parâmetros de incremento aditivo/decremento multiplicativo. A função resposta do TCP de Alta Velocidade não pode ser alcançada pelo TCP com o incremento aditivo de um segmento por tempo de percurso e com o decremento multiplicativo de cortar pela metade a janela de congestionamento corrente. É necessário modificar ambos parâmetros de incremento

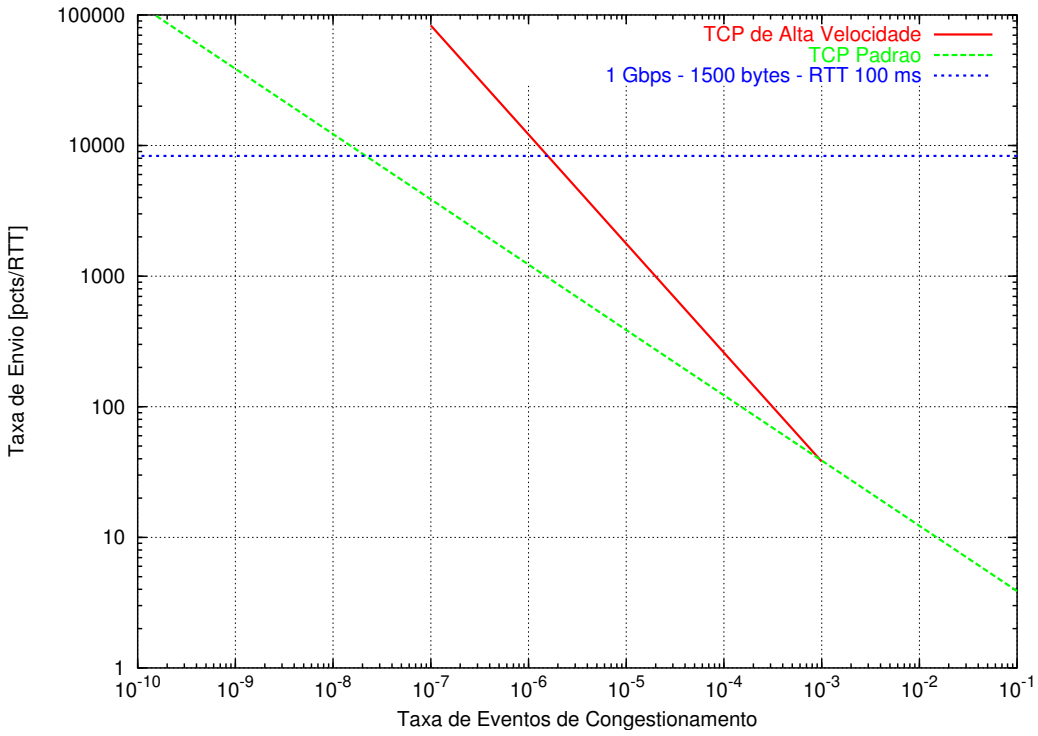


Figura 3.1: Função Resposta do TCP de Alta Velocidade

e decremento. Assim sendo, o TCP de Alta Velocidade terá de aumentar a janela de congestionamento por $a(w)$ segmentos por tempo de percurso na ausência de congestionamento e deixar a janela de congestionamento decrescer para $w*(1-b(w))$ segmentos em resposta a um tempo de percurso com um ou mais eventos de congestionamento. Na fase de Prevenção de Congestionamento, seu comportamento pode ser expresso pelas seguintes equações:

Prevenção de Congestionamento

$$\textbf{ACK} \quad : \quad CWND \leftarrow CWND + \frac{a(CWND)}{CWND} \tag{3.1}$$

$$\textbf{DROP} \quad : \quad CWND \leftarrow CWND - b(CWND) \times CWND \tag{3.2}$$

Nós podemos encontrar a expressão para as funções $a(w)$ e $b(w)$ baseado-nos nos três parâmetros definidos para o TCP de Alta Velocidade. Para $w = High\_Window$, nós temos de especificar uma taxa de eventos de congestionamento de High_P. Para o TCP Padrão, $a(w) = 1$ e $b(w) = 1/2$, independente do valor de w. O TCP de Alta Velocidade usa os mesmos valores de $a(w)$ e $b(w)$ para $w <= Low\_Window$. Estes

parâmetros podem ser expressos graficamente na Figura 3.2.

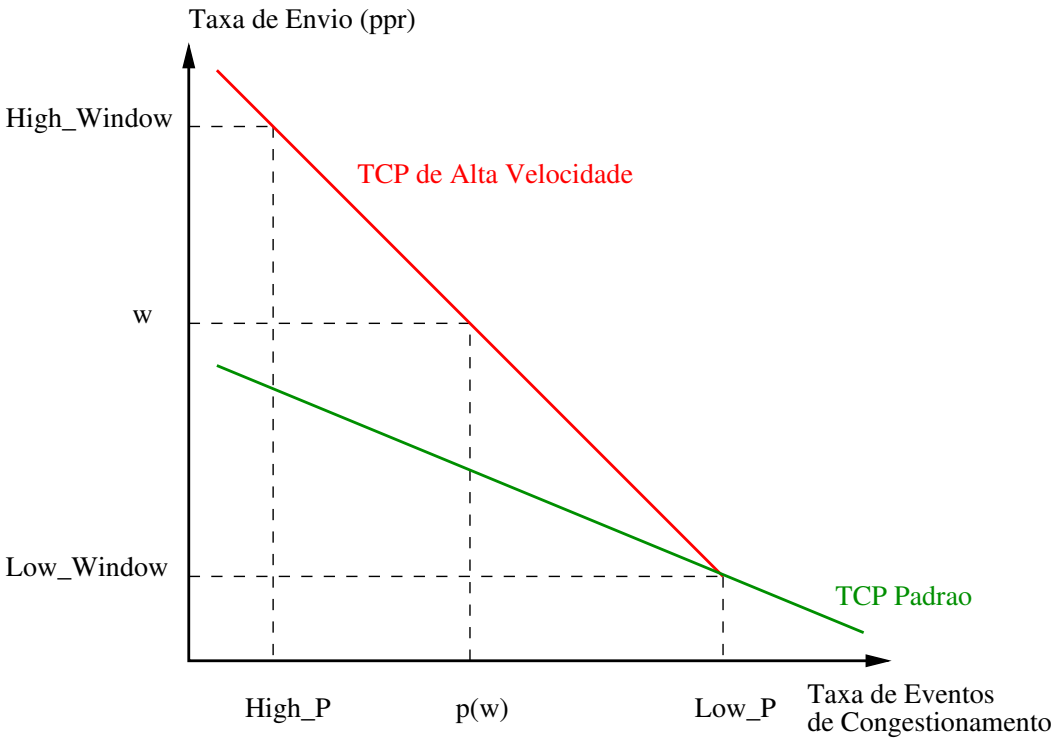


Figura 3.2: Parâmetros do TCP de Alta Velocidade em Escala Log-Log

O TCP de Alta Velocidade mantém a propriedade de ter uma linha reta para a função resposta em escala log-log (assim como o faz o TCP Padrão para congestionamento de baixo a moderado). Isto resulta, de acordo com [23], na seguinte função resposta, para valores de janela de congestionamento média W maiores do que Low_Window:

$$W = \left(\frac{p}{Low\_P}\right)^{S} \times Low\_Window \tag{3.3}$$

onde Low_P é a taxa de eventos de congestionamento correspondente a Low_Window, p é a taxa de eventos de congestionamento para a janela de congestionamento média W e S é a seguinte constante:

$$S = \frac{\log High\_Window - \log Low\_Window}{\log High\_P - \log Low\_P} \tag{3.4}$$

De [23], temos o seguinte resultado para o relacionamento entre $a(w)$ e $b(w)$:

$$a(w) = \frac{(High\_Window)^2 \times High\_P \times 2 \times b(w)}{2 - b(w)} \tag{3.5}$$

Outro parâmetro $High\_Decrease$ é usado para especificar o parâmetro de decremento $b(w)$ para $w = High\_Window$. O valor especificado em [18] é $High\_Decrease = 0.1$. Dado que o parâmetro $b(w) = 1/2$ para $w = Low\_Window$ e $b(w) = High\_Decrease$ para $w = High\_Window$, é necessário especificar os valores de $b(w)$ para os outros valores de $w > Low\_Window$. Em [23], temos $b(w)$ variando linearmente com o log de w. Portanto temos:

$$b(w) = \frac{\big(High\_Decrease - 0.5\big) \times \big(\log w - \log Low\_Window\big)}{\log High\_Window - \log Low\_Window} + 0.5 \tag{3.6}$$

$$a(w) = \frac{w^2 \times p(w) \times 2 \times b(w)}{2 - b(w)} \tag{3.7}$$

onde $p(w)$ é:

$$p(w) = \left(\frac{w}{Low\_Window}\right)^{\frac{1}{S}} \times Low\_P \tag{3.8}$$

Um exemplo do efeito desta modificação pode ser visto na Figura 3.3. O TCP de Alta Velocidade apresenta o mesmo comportamento do TCP Padrão durante a Partida Lenta. Todavia, eles possuem um comportamento distinto durante a fase de Prevenção de Congestionamento ocasionado pelo diferente algoritmo adotado pelo TCP de Alta Velocidade. Enquanto que o TCP Padrão pode apresentar um crescimento bem lento da sua janela de congestionamento, o TCP de Alta Velocidade, nas mesmas condições é capaz de abri-la mais rapidamente.

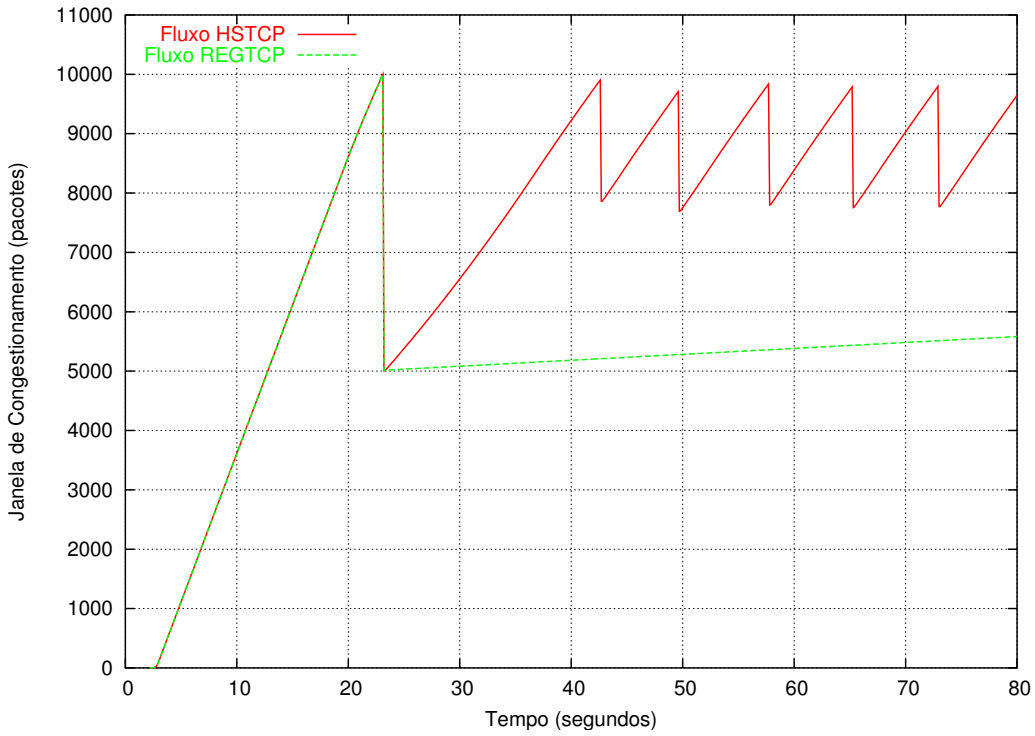


Figura 3.3: Diferença de Comportamento da Janela de Congestionamento

3.3 Seleção dos Valores dos Parâmetros

Para selecionar Low_Window é necessário escolher parâmetros conservadores que forneçam compatibilidade reversa com o TCP Padrão. Isto requer uma função resposta que seja bem próxima da função resposta do TCP Padrão para taxas de evento de congestionamento de 10^{-3}, 10^{-2} e 10^{-1}. Nós colocamos Low_Window em 38 segmentos de MSS, correspondendo a uma taxa de descarte de pacotes de 10^{-3} para o TCP Padrão. O parâmetro de decréscimo $b(w)$ para este ponto será 1/2, assim como ele o é para o TCP Padrão.

Para especificar o ponto superior da função resposta do TCP de Alta Velocidade, é necessário considerar a taxa de transferência sustentada e a taxa de descarte de pacotes esperada. Por exemplo, a janela de congestionamento média de 83.000 segmentos é aproximadamente a janela necessária para sustentar uma taxa de transferência de 10 Gbps, para uma conexão TCP com tamanho de pacote padrão de 1500 bytes e tempo de percurso de 100ms. Para o High_Window de 83.000, pode ser especificada uma High_P de 10^{-7}; isto é, com o TCP de Alta Velocidade, uma taxa de descarte de pacotes de 10^{-7}

permite conexões de TCP de Alta Velocidade alcançar uma janela de congestionamento média de 83.000 segmentos.

Estes valores colocam um alvo atingível para ambientes de alta velocidade, bem como ainda permitem uma imparcialidade aceitável para a função resposta do TCP de Alta Velocidade quando competindo com o TCP Padrão em ambientes com taxa de descarte de pacotes de 10^{-4} ou 10^{-5}.

Usando os parâmetros expressos anteriormente para a função resposta do TCP de Alta Velocidade, nós obtemos os valores descritos na Tabela 3.2.

Taxa de Perda	Janela de Congestionamento (pcts)	RTTs Entre Perdas (segs)
10^{-2}	12	8
10^{-3}	38	25
10^{-4}	263	38
10^{-5}	1795	57
10^{-6}	12279	83
10^{-7}	83981	123
10^{-8}	574356	180
10^{-9}	3928088	264
10^{-10}	26864653	388

Tabela 3.2: Função Resposta do TCP de Alta Velocidade

3.4 Imparcialidade

A imparcialidade [1] é a taxa de transmissão relativa entre os fluxos compartilhando o mesmo enlace, no qual existe uma alocação equitativa dos recursos de banda. Manter a imparcialidade entre múltiplas conexões homogêneas e heterogêneas na rede é um aspecto essencial para um protocolo ser amplamente aceito. Entretanto, o grau de imparcialidade é muito dependente da funcionalidade fornecida pela rede. A resolução de problemas de imparcialidade pode estar distribuída em funcionalidades providas pela camada de rede e pela camada de transporte.

Não é difícil prever a imparcialidade relativa entre fluxos do TCP de Alta Velocidade e do TCP Padrão. Usando as duas funções resposta e definindo que o TCP Padrão tem uma janela de congestionamento média de $W\_Standard$ e o TCP de Alta Velocidade

[1] O termo imparcialidade é usado em substituição ao termo inglês *fairness*. No contexto deste trabalho, quando a imparcialidade relativa aumenta, um protocolo é mais "parcial" do que outro, fazendo uso de uma quantidade maior da banda disponível, em relação à quantidade utilizada por outro protocolo.

tem uma janela de congestionamento média superior de $W\_HighSpeed$, a imparcialidade relativa será $W\_HighSpeed/W\_Standard$. Isto está ilustrado abaixo. Para os parâmetros escolhidos para a função resposta do TCP de Alta Velocidade, a imparcialidade relativa é descrita na Tabela 3.3.

Taxa de Descarte de Pacotes P	Imparcialidade Relativa
10^{-2}	1.0
10^{-3}	1.0
10^{-4}	2.2
10^{-5}	4.7
10^{-6}	10.2
10^{-7}	22.1
10^{-8}	47.9
10^{-9}	103.5
10^{-10}	223.9

Tabela 3.3: Imparcialidade Relativa entre as Funções Resposta do TCP de Alta Velocidade e do TCP Padrão

A Tabela 3.3 pode ser entendida deste modo: para uma taxa de descarte de pacotes de 10^{-4}, um fluxo do TCP de Alta Velocidade pode esperar ter uma taxa de transferência 2.2 vezes maior que um fluxo do TCP Padrão, dadas as mesmas condições de tempo de percurso e tamanho de pacote.

Capítulo 4

Proposta de Avaliação do TCP de Alta Velocidade

A proposta geral deste trabalho foi estudar a eficácia do TCP de Alta Velocidade em enlaces de alta velocidade e longa distância, como um mecanismo para transferência de grande volume de dados, enquanto mantendo imparcialidade com outros tipos de TCP já em uso.

Para cumprir este objetivo geral, este estudo teve questões específicas para responder, descritas a seguir:

1. Qual é comportamento do TCP de Alta Velocidade em situações nas quais o TCP Padrão possui baixo desempenho?

2. É possível utilizar o TCP de Alta Velocidade junto com o TCP Padrão e manter uma imparcialidade aceitável?

3. Qual é o efeito da política de enfileiramento do roteador no desempenho do TCP de Alta Velocidade e na imparcialidade entre o TCP de Alta Velocidade e o TCP Padrão?

4. O TCP de Alta Velocidade pode ser um substituto para outros tipos de transferência volumosa de dados?

4.1 Seleção da Abordagem

Existem várias possíveis abordagens para o desenvolvimento desta investigação. A primeira abordagem seria encontrar uma solução analítica para cada questão. Muito embora este método forneça uma resposta precisa, na maioria das vezes é muito difícil formular uma solução ampla que cubra todos os aspectos da investigação.

A segunda abordagem seria a implementação de um protótipo, sua utilização na Internet nas situações em que esperamos que ele lide e a coleta de dados de sua utilização. Seria então possível verificar o emprego num ambiente real com este método, bem como um objetivo excelente para aplicações de longa distância na Internet. Entretanto, os custos e as dificuldades de avaliar sistemas diretamente através deste caminho seriam proibitivos [24].

A terceira abordagem seria usar um emulador do protocolo em uma rede de teste controlada. Emulação é uma ferramenta interessante que oferece muitas das vantagens da avaliação direta na Internet, mas não elimina o problema de obtenção de recursos computacionais remotos e acesso de rede [60]. A emulação permite que máquinas rodando o serviço atual em uma rede local experimentem atrasos e limitações de banda normalmente impostos por redes de longa distância. Este método também requereria que o sistema fosse implementado pelo menos num estágio de protótipo e também forçaria que os experimentos não executassem mais rápido que o tempo real.

A próxima abordagem possível para esta investigação seria o uso de simulação. Este método evita muitos dos custos existentes nos métodos anteriores e é indicado como o método para a primeira análise de condições e comportamentos de rede complexos [6]. Usando este método, é possível avaliar protocolos de rede sobre várias condições de rede e investigar interações entre protocolos não previstas.

Como este presente trabalho é um dos primeiros experimentos do TCP de Alta Velocidade, é razoável se ter o método de simulação como a abordagem escolhida. A outra razão para esta escolha foi a disponibilidade de bons simuladores de rede de propósito geral que são amplamente aceitos pela comunidade científica da área de redes e que possuíam as características requeridas para o desenvolvimento deste estudo.

4.2 Delimitação do Escopo

Muito embora o uso de simulação permita a investigação de um protocolo em uma rica variedade de situações, nem todas as condições de rede foram de interesse. Nós limitamos os cenários de investigação a casos de interesse selecionados. O foco principal foi o comportamento do TCP de Alta Velocidade e do TCP Padrão em situações em que ambos estivessem em estado estacionário ou perto do estado estacionário (quando a janela de congestionamento oscila em torno de seu ponto de equilíbrio, caracterizado por uma forma de onda em dente de serra). A fase de prevenção de congestionamento do TCP foi de particular interesse, porque é onde o algoritmo AIMD trabalha, conseqüentemente é onde o algoritmo do TCP de Alta Velocidade atua. Estados transitórios foram também de interesse, mas somente quando eles fizessem diferença no comportamento e desempenho do estado estacionário.

Algumas considerações também precisam ser feitas para a fase de Partida Lenta, devido a sua influência no desempenho geral. O algoritmo usado na Partida Lenta tem um considerável impacto na fase de Prevenção de Congestionamento. Quando operando com grandes janelas de congestionamento, é possível de se ter milhares de pacotes descartados de uma única janela, o que faz com que a recuperação de uma conexão TCP seja muito lenta.

Fluxos TCP de longa duração são de maior interesse para enlaces de alta velocidade e longa distância, quando uma grande quantidade de dados precisa ser transmitida. Portanto, como consideração geral, a maioria dos fluxos usados neste trabalho tiveram longa duração, no qual o estado estacionário era bem maior que o estado transiente.

Esta investigação foi desenvolvida também usando-se um cenário com topologia simples para evitar interações mais complexas e reduzir o número de variáveis a coletar e estudar.

O TCP *SACK* [13] foi utilizado neste trabalho para a comparação do controle de congestionamento do TCP Padrão com o TCP de Alta Velocidade, porque ele possui o melhor desempenho para estas situações quando comparado com o *Tahoe*, *Reno* e *Newreno*. O TCP SACK permite que o receptor de uma conexão TCP informe ao transmissor sobre múltiplos pacotes descartados dentro de uma única janela de dados. O TCP *Vegas* não foi incluído neste estudo devido ao seu reduzido desempenho quando

competindo com outras versões do TCP [37], no período em que este estudo foi desenvolvido.

O gerenciamento de enfileiramento no roteador foi restrito à técnica de descarte *DropTail* (DT) e *Random Early Detection* (RED). DT é a técnica tradicional para o gerenciamento do tamanho de fila do roteador e RED é recomendado como o melhor mecanismo padrão para o gerenciamento ativo de fila (*Active Queue Management* ou AQM) [4]. No caso do RED, o *Explicit Congestion Notification* (ECN) [52] foi empregado.

O DT controla o comprimento da fila usando um esquema FIFO (*First In First Out*). Neste esquema, cada novo pacote que chega na porta de entrada da fila é descartado quando o espaço de *buffer* da fila está cheio. Por seu lado, o RED descarta probabilisticamente um pacote que chegue à fila, mesmo que a fila da interface de saída não esteja cheia.

4.3 Metodologia

Nesta seção é apresentada a metodologia usada durante esta pesquisa. É apresentada a ferramenta de simulação utilizada, o ambiente de rede usado, os tipos de fluxos TCP empregados e como os dados foram coletados. Ao final, são descritas as métricas e os cenários de rede usados na avaliação.

Seleção do Simulador

O simulador NS-2 (*Network Simulator - version 2*) [50] foi utilizado nos experimentos pelas seguintes razões:

- é amplamente utilizado pela comunidade de pequisa da área de redes;
- é um simulador de pacotes que fornece uma rica biblioteca de componentes de rede e protocolos;
- possui uma boa escalabilidade que permite ser utilizado em diferentes cenários e com diferentes números de fluxos e nós;

- é um software de código aberto, podendo ser modificado se necessário; e
- já implementa o TCP de Alta Velocidade.

O NS-2 possui características adicionais que facilitam este trabalho: os nós finais e enlaces possuem parâmetros, tais como banda e atraso, que permitem o seu ajuste para diferentes cenários; os roteadores têm implementadas as políticas de gerenciamentos de filas DT e RED; existem capacidades de monitoramento para rastrear os pacotes nas filas dos roteadores bem como nos fluxos TCP; o simulador possui suporte matemático incluindo gerador de números randômicos e funções de distribuição. Os experimentos de simulação foram escritos para o NS-2 em TCL [59], uma rica linguagem de *scripts*.

4.3.1 Política de Enfileiramento do Roteador

Neste estudo, nós verificamos as diferenças provocadas pelo uso de dois esquemas de gerenciamento de enfileiramento: DT e RED (o ECN está ativo quando o RED foi empregado). Muito embora não exista a intenção de realizar um estudo profundo das questões de gerenciamento de enfileiramento do roteador, sua influência é forte o suficiente para garantir alguma atenção. Estas políticas de enfileiramento foram escolhidas por serem amplamente conhecidas, pesquisadas e empregadas.

Buffers de roteadores são um elemento essencial para uma rede por comutação de pacotes. Eles absorvem as chegadas de pacote em rajada e reduzem o potencial de perdas. Quanto maior for o *buffer*, maior será a capacidade de absorver grandes rajadas. Todavia, isto faz crescer a carga e aumenta os atrasos de enfileiramento.

O DT é o meio mais simples de realizar o gerenciamento de fila em um roteador. Ele controla o comprimento da fila usando um esquema FIFO (*First In First Out*). Neste esquema, cada novo pacote que chega na porta de entrada da fila é descartado quando o espaço de *buffer* da fila está cheio. Van Jacobson propôs em 1988 que o espaço de *buffer* na fila deveria ser não menos que o produto banda atraso. Também foi proposto que o atraso deveria ser a média do tempo de percurso fim-a-fim de todos os fluxos compartilhando o enlace gargalo [47].

O DT possui uma tendência contra fluxos em rajada, porque um fluxo em rajada tende a ter múltiplos pacotes chegando na fila aproximadamente ao mesmo tempo. Portanto, o roteador irá descartar vários pacotes do mesmo fluxo ao mesmo tempo. O DT também produz sincronização global, porque pacotes de todos os fluxos são descartados quando a fila está cheia, conduzindo todos os fluxos a retraírem-se ao mesmo tempo. Todos os fluxos seguem o mesmo algoritmo para a taxa de aumento, seguindo para a mesma situação anterior, aproximadamente ao mesmo tempo [4]. *Lockouts* são outro problema com o DT. O DT pode permitir que uma única conexão controle o enlace gargalo contribuindo para o aumento da imparcialidade, mas não necessariamente sendo traduzido numa utilização de baixa do enlace [49].

Usando o RED [4, 22], um roteador irá descartar probabilisticamente um pacote que chegue à fila, mesmo que a fila da interface de saída não esteja cheia. A razão para este descarte precoce vem do fato que a perda de pacote é o indicador primário de congestionamento para uma conexão TCP. Descartando pacotes antes que a fila do roteador esteja completamente cheia, as conexões TCP compartilhando da fila irão reduzir suas taxas de transmissão e assegurar que a fila não transborde. Outra conseqüência deste descarte antecipado é que o RED força a imparcialidade, porque a fração dos pacotes descartados de cada conexão é aproximadamente proporcional à porção de banda da conexão.

O algoritmo RED usa uma média ponderada do comprimento total da fila para determinar quando descartar pacotes. Quando um pacote chega à fila, se a média ponderada do comprimento da fila é menor que um valor limite mínimo (min_{th}), então nenhuma ação de descarte será tomada e o pacote será simplesmente enfileirado. Se a média é maior que o valor limite mínimo, porém menor que limite máximo (max_{th}), um teste de descarte antecipado é feito conforme descrito abaixo, até a taxa descarte máxima (max_p), quando o tamanho médio da fila alcança max_{th}. Um tamanho médio de fila na faixa entre os limites indica que algum congestionamento começou e os fluxos devem ser notificados através de descarte de pacotes. Se a média do comprimento da fila é maior que o valor do limite máximo, uma operação forçada de descarte irá ocorrer. Um comprimento de fila médio nesta faixa indica um congestionamento persistente e pacotes precisam ser descartados para evitar que a fila fique cheia persistentemente. Um descarte forçado também é usado quando a fila está cheia mas o tamanho médio de fila ainda está abaixo do limite máximo.

Usando a média ponderada, o RED evita uma reação excessiva às rajadas e reage a tendências de longo prazo. Além disto, devido aos limites serem comparados com a média ponderada, é possível que nenhum descarte forçado ocorra quando o comprimento instantâneo da fila for bem longo. Uma representação gráfica dos parâmetros do RED é dada na Figura 4.1.

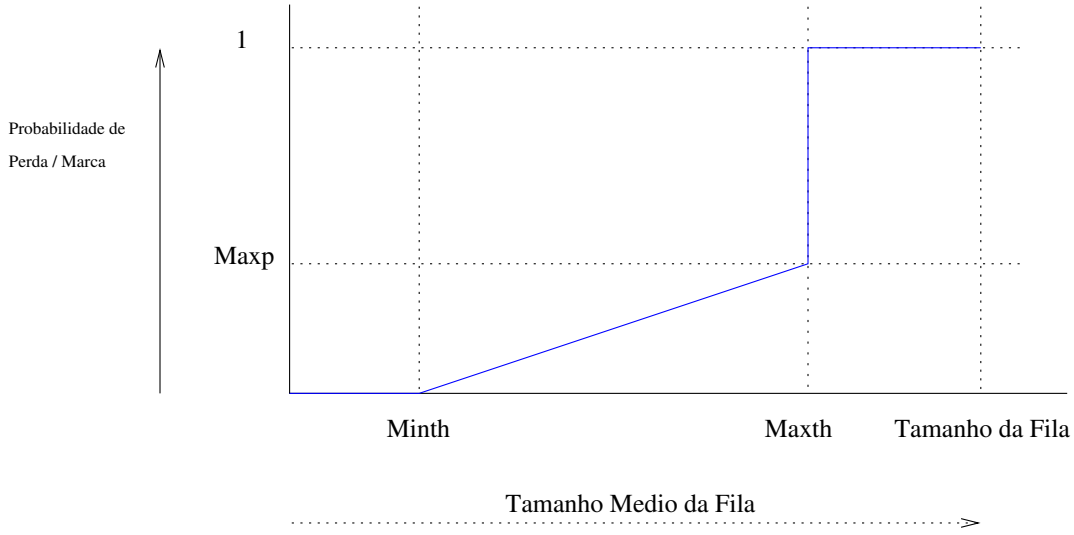


Figura 4.1: Parâmetros do RED

Adaptative RED [20] é uma variação do RED a qual retém a estrutura básica do RED e ajusta dinamicamente o parâmetro max_p para manter o tamanho de fila médio entre os limites mínimo e máximo. O objetivo do *Adaptative RED* é reduzir a taxa de perdas de pacote e a variação no atraso de enfileiramento.

O modo suave do RED [16] modifica a função de descarte do RED para o caso quando o tamanho médio de fila excede max_{th}. A probabilidade de descarte aumenta linearmente entre max_{th} e o tamanho total do *buffer* com uma inclinação de $(1-max_p)/max_p$. Obviamente, o RED tem que descartar um pacote que chega se o tamanho instantâneo da fila é igual ao tamanho total do *buffer*.

Outra extensão do RED é *marcar* o cabeçalho do IP ao invés de descartar pacotes, quando o tamanho médio de fila está entre min_{th} e max_{th}, ou entre min_{th} e o tamanho do *buffer* (quando *Adaptative RED* e o modo suave são empregados conjuntamente). Os sistemas na ponta da conexão devem cooperativamente usar esta marca no IP como um sinal de que a rede está congestionada e diminuir a taxa de transferência. Isto é conhecido com Notificação Explícita de Congestionamento (*Explicit Congestion Notification* ou

ECN) [53].

O ECN objetiva prover o TCP de um mecanismo alternativo para a detecção de congestionamentos incipientes na rede. Isto é, o transmissor TCP que tem suporte ao ECN não precisa depender somente do descarte de pacotes para detectar congestionamento e reduzir sua taxa de envio.

O ECN requer suporte tanto dos roteadores quanto dos computadores nas pontas da conexão. Os computadores negociam a capacidade ECN durante o estabelecimento da conexão TCP. Se ambos os lados estiverem habilitados a trabalhar com ECN, o TCP transmissor indica isto acionando um bit em cada pacote enviado. Roteadores habilitados para ECN são responsáveis por monitorar os níveis de congestionamento e marcar os pacotes das fontes habilitadas para ECN quando o congestionamento se torna crítico, ao invés de esperar passivamente que o *buffer* fique sem espaço e tenha de recorrer ao descarte de pacotes. O ECN conta com a habilidade do roteador detectar congestionamento incipientes, ao contrário do DT. Portanto o roteador precisa usar algum mecanismo de Gerenciamento Ativo de Filas (*Active Queue Management* ou AQM), tal como o empregado pelo RED.

O *Adaptative RED*, com o modo suave e suporte para ECN foram usados para todas as simulações neste trabalho que foram conduzidas com o gerenciamento de enfileiramento de roteador RED.

4.3.2 Ambiente de Simulação

Esta seção descreve o ambiente usado para o desenvolvimento deste trabalho. Todas as condições e parâmetros usados nos enlaces de rede, nós e fluxos são apresentadas.

Topologia de Rede

A topologia de rede escolhida para as simulações foi a bem conhecida *barra de pesos* (*dumbbell*) [62] com um único gargalo, como apresentado na Figura 4.2. Esta topologia fornece uma excelente plataforma para estudar os efeitos das interações entre

os fluxos, pois ela representa o cenário de uma típica conexão TCP de longa distância, no qual existe um único enlace ótico ligado dois pontos distantes. Todo o tráfego da simulação passa através do enlace gargalo e todos os pontos finais estão conectados a ele. Nós definimos como *direta* a direção do nó N1 para o nó N2 e *reversa* como sendo a direção do nó N2 para o nó N1.

O enlace gargalo foi o enlace principal. A não ser que seja diferentemente dito, a banda do enlace principal foi 1 Gbits/seg, o seu atraso foi de 50 ms e o tipo de gerenciamento de enfileiramento do roteador foi RED ou DT.

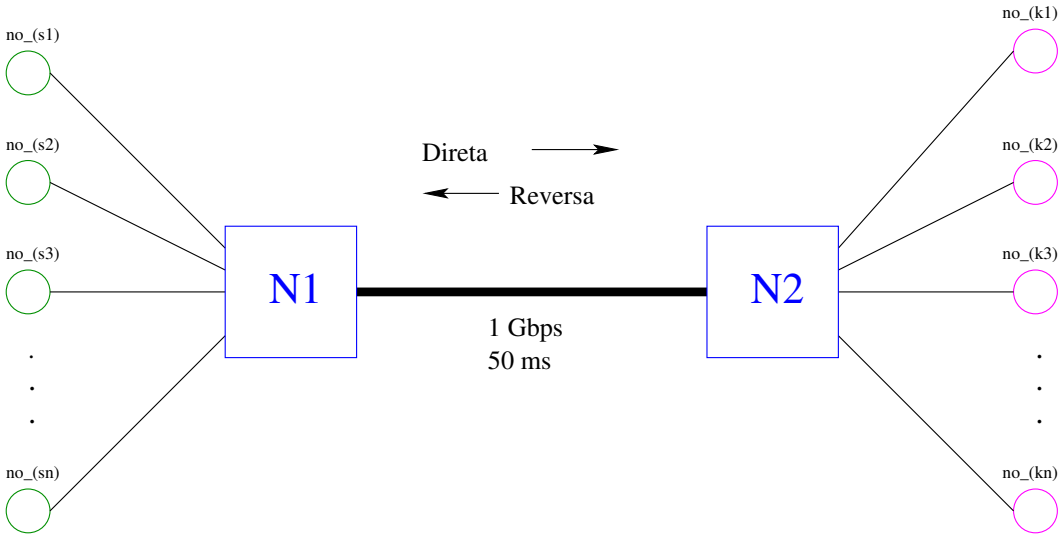


Figura 4.2: Topologia de Rede

Nós utilizamos dois tipos de gerenciamento de enfileiramento no roteador, DT e RED. O tamanho de fila usado foi BDP em pacotes. O valor padrão foi de 8.333 pacotes, porque este valor coincide com o BDP para um enlace de 1 Gbps, com RTT de 100ms e tamanho de pacote de 1500 bytes. Os demais parâmetros para o RED estão descritos na Tabela 4.1. Uma explicação mais detalhada do funcionamento do RED encontra-se na seção 6.2.3, na página 104.

Parâmetros RED	Valor	Significado
gentle_	true	modo gentil do RED ativo
adaptive_	1	RED adaptativo ativo
bottom_	0.0001	valor mínimo para max_p
setbit_	1	marcação de ECN ativa
targetdelay_	0.005	tamanho alvo da fila

Tabela 4.1: Parâmetros RED

O Modelo de Erros (*Error Model*), definido no NS-2, foi usado para a simulação de perdas no enlace. Ele simula erros e perdas no nível de enlace ou marcando a bandeira (*flag*) de erro do pacote ou enviando o pacote para descarte. Neste trabalho, a unidade do Modelo de Erros foi configurada em unidade de *pacotes* e a variável randômica para ser, por simplicidade, uniformemente distribuída entre 0 e 1. A *taxa de erros* é um parâmetro baseado no nível de erros de enlace desejado. Este Modelo de Erros foi somente utilizado no enlace gargalo. A não ser quando indicado num experimento com perdas no enlace gargalo, o modelo de erros ficou desativado.

Os nós finais são os pontos finais das conexões TCP. Eles estão ligados com os pontos finais do enlace gargalo, N1 e N2. Metade dos nós finais estão conectados com N1 e a outra metade com N2. Cada um destes enlaces teve a banda configurada para 100 Gbps para assegurar que eles não limitariam os fluxos TCP. Eles usaram a política de enfileiramento do roteador DT e um atraso de enlace entre 1 e 73 milisegundos (muito abaixo do enlace gargalo). Cada nó possui um atraso levemente diferente para promover diversidade entre os comportamentos dos fluxos.

Configuração dos Fluxos TCP

Os fluxos nos experimentos usaram TCP. Estes fluxos tiveram um conjunto de parâmetros em comum. Eles tiveram o bit ECN acionado para reagir a pacotes marcados com ECN vindos do gerenciamento de enfileiramento do roteador [51, 52]; o tamanho de pacote foi de 1500 bytes, porque este é o tamanho de mais da metade do volume em bytes do tráfego em conexões de longa distância atualmente [44]; o tamanho máximo de janela foi grande o suficiente para não impor limites pelo tamanho máximo da janela de anúncio do receptor; foram configurados tempos randômicos entre os envios para evitar efeitos de fase (*phase effects*) [21]; caso não seja especificado diferentemente, os fluxos utilizaram a versão modificada do algoritmo de Partida Lenta para grandes janelas de congestionamento (*Limited Slow-Start*) para evitar uma perda volumosa de pacotes nesta fase [17]; e o tamanho mínimo de cabeçalho TCP foi configurado, com nenhum cabeçalho opcional. O conjunto completo dos parâmetros TCP está detalhado na Tabela 4.2.

Parâmetros TCP	Valor	Unidade	Significado
ecn_	1	ativo / não ativo	Ativação do ECN
window_	100000	pacotes	Tamanho Máximo de Janela
packetSize_	1500	bytes	Tamanho do Pacote
overhead_	0.000008	adimensional	Adição de Tempo Randômico entre Envios
max_ssthresh_	100	pacotes	Valor do Limiar da Partida Lenta Limitada

Tabela 4.2: Parâmetros TCP

O agente TCP usado no transmissor e receptor foi o SACK1. O TCP/SACK1 implementa o protocolo de transporte BSD TCP Reno com extensão do Reconhecimento Seletivo descrito em [13]. Este tipo de TCP implementa a maioria das melhorias do TCP disponíveis atualmente. O HSTCP também está baseado na implementação deste tipo de TCP, mas com as modificações no algoritmo de controle de congestionamento. O FTP foi a aplicação usada para transmitir os dados através das conexões TCP. Ele simula uma transferência de grande volume de dados entre dois nós.

O HSTCP foi implementando no NS-2 como uma opção para o controle da janela de congestionamento. Múltiplos fluxos HSTCP não iniciavam ao mesmo tempo. Ao contrário, eles começavam aleatoriamente no primeiro décimo do tempo total de simulação. A direção *direta* foi usada para todos os fluxos HSTCP. Os valores dos parâmetros do HSTCP usados estão descritos na Tabela 4.3 e foram retirados dos valores propostos em [18].

Parâmetros do HSTCP	Valor	Unidade
Low_Window_	31	pacotes
High_Window_	83000	pacotes
High_P_	0.0000001	adimensional
High_Decrease_	0.1	adimesnional

Tabela 4.3: Parâmetros do HSTCP

Para comparação, o HSTCP foi executado em conjunto com a implementação do TCP Padrão (referido também neste trabalho como TCP Regular ou REGTCP). Estes dois tipos de fluxo utilizaram a implementação do TCP/SACK1 como explicado anteriormente. Os mesmos parâmetros básicos foram usados para todos os fluxos TCP e também estes fluxos iniciavam randomicamente no primeiro décimo do tempo total de simulação para evitar efeitos de fase. Estes fluxos usaram somente a direção *direta* para transmitir seus dados.

Desde que o tráfego dominante atualmente na Internet é proveniente de transações *web*, um tipo adicional de fluxo foi incorporado aos experimentos para reduzir a regularidade do tráfego e obter resultados mais realistas. O uso deste tipo de fluxo teve o impacto de adicionar tráfego similar a *web* nas simulações. O módulo *PagePool/WebTraf* do NS-2 foi usado para fornecer este tipo de comportamento de tráfego. Com o uso deste módulo foi possível configurar vários parâmetros, definindo as características de servidores e clientes *web* (por exemplo, o número total de páginas por sessão, distribuição de tamanhos de página, distribuição dos intervalos entre chegada das páginas e distribuição do tamanho dos objetos). Neste trabalho, o tamanho médio dos objetos foi 10, o *pool* de servidores e clientes *web* foi configurado em 10. Dois conjuntos de *pools* foram configurados, um na direção *direta* (servidores *web* no lado do nó N1 e os clientes no outro lado do nó N2) e o outro *pool* na direção *reversa* (servidores *web* no lado do nó N2 e clientes *web* no lado de N1). Os servidores e clientes *web* foram ligados aos nós do enlace gargalo através de um enlace de 100 Gbps e um atraso de enlace variável.

Um conjunto de 20 pequenos fluxos TCP também foi usado nas simulações, 10 fluxos na direção *direta* e 10 fluxos na direção *reversa*. Estes fluxos possuíam um tamanho máximo de janela de 8 pacotes. A fonte e o destino deste fluxos foram distribuídos aleatoriamente entre os nós ligados ao enlace gargalo (nos mesmos nós usados para os fluxos HSTCP e REGTCP), de forma que eles pudessem interferir nos fluxos HSTCP e REGTCP. Estes pequenos fluxos TCP iniciavam em um tempo aleatório no primeiro terço da simulação e terminavam também em um tempo aleatório no terço final da simulação. Eles representaram as pequenas conexões de curta duração na Internet.

Os dois tipos anteriores de fluxo descritos acima constituíram o *ruído de fundo* para todas as simulações. Eles foram usados para evitar que se tivessem padrões muito regulares nos experimentos. Eles usaram menos de 1% da capacidade de enlace total disponível. Eles foram usados para quebrar a regularidade na transmissão dos pacotes nos fluxos TCP, especialmente em ambientes de ausência de interferência externa.

Outro tipo de fluxo TCP foi usado para representar tráfego em rajada. Ele foi constituído de fluxos de curta duração que se mantinham por cerca de 1,6 segundos. Este tempo permitia que estes fluxos rodassem apenas na fase de Partida Lenta durante seu tempo de vida para as condições dos experimentos, apresentando, portanto, apenas o comportamento exponencial característico desta fase do controle de congestionamento. A versão modificada do algoritmo de Partida Lenta para grandes janelas de

congestionamento não foi usada neste caso; ao contrário, o algoritmo padrão foi executado. Quando diversos fluxos em rajada eram usados em uma simulação, o seu tempo de início era randomicamente distribuído, com distribuição uniforme, durante todo o período de simulação.

Configuração da Coleta de Dados

O NS-2 fornece um módulo monitor para coletar dados. Dois monitores foram usados neste trabalho, um para monitorar a fila no enlace gargalo, rastreando as estatísticas de chegada, partida e descarte. O outro foi usado para monitorar as estatísticas por fluxo, tais como chegadas, partidas e descarte de pacotes e bytes de cada fluxo que cruzava o enlace gargalo.

Em adição a estas estatísticas no nível de rede, os monitores também provêem a capacidade de coletar estatísticas no nível de transporte. Foi possível verificar como as variáveis internas do TCP estavam se comportando. A cada *tick* de simulação, informações tais como o tamanho corrente da janela de congestionamento, o maior número de seqüência enviado, o número de respostas ECN e o valor do limiar da Partida Lenta estavam disponíveis.

Os dados agregados foram coletados duas vezes, uma após a metade do tempo de simulação ter passado e outra ao final da simulação. Com os dois conjuntos de dados, foi possível calcular os resultados para a segunda metade da simulação. Apenas a segunda metade da simulação era de interesse porque esta pesquisa estava focalizada no comportamento em estado estacionário.

Diversas variáveis foram coletadas a cada 0,1 segundos. Isto foi necessário para entender seu comportamento em tempos intermediários durante a simulação. Nas simulações foi usado o gerador de número randômico do NS-2, com sementes alimentadas heuristicamente.

Cada experimento foi simulado dez vezes, por trezentos segundos. O resultado final apresentado foi a mediana destas simulações. A mediana foi utilizada por ser menos sensível à presença de *outliers* nos resultados do que a média. A versão do NS-2 utilizada foi a NS-2.1b9a ou atualizações mais recentes disponíveis no repositório do NS-2.

4.3.3 Métricas Utilizadas na Avaliação de Desempenho

Os parâmetros coletados nesta pesquisa estão listados e descritos na Tabela 4.4.

Tipo	Descrição	Unidade
tamanho de fila	tamanho instantâneo da fila	pacotes
tamanho da janela	tamanho de janela de congestionamento corrente	pacotes
número de seqüencia	maior pacote de reconhecimento visto pelo receptor	—
pacotes descartados	nr total de pacotes descartados	pacotes
pacotes marcados	nr total de pacotes marcados com ECN	pacotes
pacotes enviados	nr total de pacotes enviados por um fluxo	pacotes
partidas de pacotes	nr total de pacotes que partiram (não descartados) de uma fila	pacotes
chegadas de pacotes	total de pacotes que chegaram na fila	pacotes

Tabela 4.4: Estatísticas Coletadas pelo Monitor

Um conjunto de métricas foi usado durante este trabalho para avaliar o desempenho do HSTCP, para medir o seu impacto em outros tipos de tráfego TCP e verificar seu comportamento em diferentes condições de rede. Antes de apresentar a formulação das métricas, algumas definições precisam ser introduzidas. Elas são apresentadas a seguir.

Banda B

Descrição: O número de bits por segundo que um enlace é projetado para transmitir.

Unidade: bits/seg

Tamanho de Pacote pct

Descrição: O tamanho do pacote na rede. Por razões de simplicidade, este tamanho é fixo em 1500 bytes.

Unidade: bytes

Intervalo de Tempo T

Descrição: Um período de tempo.

Unidade: segundos

Pacotes Enviados $PE_f(T)$

Descrição: O número de pacotes que um fluxo f transmite durante o intervalo de tempo T.

Unidade: pacotes

Pacotes Enviados Agregados $PEA_p(T)$

Descrição: O número de pacotes enviados por todos os fluxos do mesmo protocolo p durante o intervalo de tempo T, onde F é o número de fluxos pertencentes ao protocolo p .

Unidade: pacotes

Expressão:

$$PEA_p(T) = \sum_{k=1}^{F} PE_k(T)$$

Pacotes Enviados Geral $PEG(T)$

Descrição: O número de pacotes enviados por todos os protocolos pertencentes ao mesmo experimento durante o intervalo de tempo T, onde P é o número de protocolos presentes.

Unidade: pacotes

Expressão:

$$PEG(T) = \sum_{k=1}^{P} PEA_k(T)$$

Capacidade do Enlace $C(T)$

Descrição: O número de pacotes que podem ser transmitidos através de um enlace durante o intervalo de tempo T.

Unidade: pacotes

Expressão:

$$C(T) = \frac{B}{pct \times 8} \times T$$

Pacotes Descartados $PD_f(T)$

Descrição: O número de pacotes descartados do fluxo f durante o intervalo de tempo T.

Unidade: pacotes

Pacotes com ECN Marcado $ECNP_f(T)$

Descrição: O número de pacotes com ECN marcado no fluxo f durante o intervalo de tempo T.

Unidade: pacotes

Eventos de Congestionamento por Fluxo $ECF_f(T)$

Descrição: A soma dos pacotes descartados mais os pacotes com ECN marcado no fluxo f durante o intervalo de tempo T.

Unidade: pacotes

Expressão:

$$ECF_f(T) = PD_f(T) + ECNP_f(T)$$

Eventos de Congestionamento $EC_p(T)$

Descrição: A soma do número de pacotes descartados mais os pacotes com ECN marcado de todos os fluxos pertencentes ao mesmo protocolo p durante o intervalo de tempo T, onde F_p é o número de fluxos do protocolo.

Unidade: pacotes

Expressão:

$$EC_p(T) = \sum_{k=1}^{F_p} ECF_k(T)$$

Eventos de Congestionamento Geral $ECG(T)$

Descrição: A soma dos eventos de congestionamento agregados de todos os protocolos pertencentes ao mesmo experimento durante o intervalo de tempo T, onde P é o número de protocolos presentes.

Unidade: pacotes

Expressão:

$$ECG(T) = \sum_{k=1}^{P} EC_k(T)$$

Métricas Gerais

Taxa de Eventos de Congestionamento $TEC_p(T)$

Descrição: A razão entre os eventos de congestionamento $EC_p(T)$ e o número de pacotes enviados agregados do mesmo tipo de protocolo p durante o intervalo de tempo T.

Unidade: adimensional

Expressão:

$$TEC_p(T) = \frac{EC_p(T)}{PEA_p(T)}$$

Taxa de Eventos de Congestionamento Geral $TECG(T)$

Descrição: A razão entre o número de eventos de congestionamento geral e o número de pacotes enviados por todos os protocolos pertencentes ao mesmo experimento durante o intervalo de tempo T.

Unidade: adimensional

Expressão:

$$TECG(T) = \frac{ECG(T)}{PEG(T)}$$

Métricas de Utilização do Enlace

Estas métricas são usadas para verificar quanto de banda no enlace gargalo é usado por um fluxo específico ou por um dado protocolo. Parte das métricas usadas nesta pesquisa vieram das definições usadas em [61] e [8].

Utilização do Enlace por Fluxo $UEF_f(T)$

Descrição: A fração do enlace gargalo utilizado por um fluxo f, durante o intervalo de tempo T em um enlace com Capacidade de Enlace $C(T)$.

Unidade: adimensional

Expressão:

$$UEF_f(T) = \frac{PE_f(T)}{C(T)} \times 100\%$$

Utilização do Enlace $UE_p(T)$

Descrição: A soma da utilização do enlace por fluxo de todos os fluxos pertencentes ao mesmo protocolo p durante o intervalo de tempo T, onde F_p é o número de fluxos de um protocolo p.

Unidade: adimensional

Expressão:

$$UE_p(T) = \sum_{k=1}^{F_p} UEF_k(T)$$

Banda Roubada $BR_{p_1,p_2}(T)$

Descrição: A diferença entre a utilização do enlace alcançada por N fluxos pertencetes a um protocolo p_1 quando eles estão competindo contra M fluxos pertencentes ao mesmo protocolo p_1 e a utilização do enlace atingida pelo mesmos N fluxos pertencentes ao protocolo p_1 quando eles estão competindo contra M fluxos pertencentes ao protocolo p_2, durante o intervalo de tempo T.

Unidade: adimensional

Métricas de Imparcialidade

Estas métricas foram usadas para verificar o impacto do uso de um protocolo diferente (no caso HSTCP) em outros protocolos já existentes (no caso REGTCP).

Imparcialidade Relativa por Fluxo IRF_{f_1,f_2}

Descrição: Avalia a razão entre a utilização do enlace por um fluxo f_1 pertencente ao protocolo p_1 e a utilização do enlace por um fluxo f_2 pertencente ao protocolo p_2, durante o intervalo de tempo T.

Unidade: adimensional

Expressão:

$$IRF_{f_1,f_2} = \frac{UEF_{f_1}(T)}{UEF_{f_2}(T)}$$

Imparcialidade Relativa IR_{p_1,p_2}

Descrição: Avalia a razão entre a utilização do enlace por um conjunto de fluxos pertencente ao protocolo p_1 e a utilização do enlace por um conjunto de fluxos pertencente ao protocolo p_2, durante o intervalo de tempo T.

Unidade: adimensional

Expressão:

$$IR_{p_1,p_2} = \frac{UE_{p_1}(T)}{UE_{p_2}(T)}$$

4.3.4 Descrição dos Cenários dos Experimentos

Utilizamos três conjuntos de fluxos na maior parte deste estudo. O primeiro conjunto continha apenas fluxos HSTCP, o segundo era composto somente de fluxos REGTCP e o terceiro conjunto possuía ambos, fluxos HSTCP e REGTCP. O primeiro e segundo conjuntos permitiram a comparação entre os fluxos REGTCP e HSTCP. O terceiro conjunto de fluxos permitiu-nos observar a interação entre os fluxos REGTCP e HSTCP. O número de fluxos em cada conjunto variou em cada experimento.

Estes três conjuntos de fluxos foram expostos a diferentes condições de rede. Estas diferentes condições de rede permitiram-nos ver a variação das métricas e produzir um quadro geral do comportamento do HSTCP e sua interação com o REGTCP.

No primeiro ambiente de rede não havia outra fonte de tráfego e interferência além da gerada pelos fluxos REGTCP e HSTCP. Este ambiente de rede é referenciado como *Condição Ideal*. Esta situação é interessante porque foi possível estudar o desempenho dos conjuntos de fluxo sem interferência externa. Ele serviu de base para comparar o desempenho dos conjuntos de fluxos em outros ambientes de rede.

O segundo ambiente de rede representou a situação onde haviam perdas sistêmicas (ou perdas não diretamente relacionadas a congestionamento). Ele foi chamado de *Condição de Enlace com Perdas*. Um número de pacotes era randomicamente descartado dos fluxos com uma distribuição uniforme, usando-se o Modelo de Erros do simulador NS-2 e com uma taxa de descarte definida. Isto nos permitiu analisar os efeitos de diferentes níveis de enlace com perdas nos conjuntos de fluxos.

O terceiro ambiente de rede explorou a reação dos três conjuntos de fluxos a tráfego em rajada, portanto foi chamado de *Condição de Tráfego com Característica de Rajada*. O tráfego com característica de rajada era composto de fluxos de TCP Padrão de curta duração, atuando por poucos segundos e somente durante a fase de Partida Lenta. Desta forma foi possível verificar o quanto eles interferiam no comportamento dos três conjuntos de fluxos e, conseqüentemente, em suas métricas.

Capítulo 5

Resultados dos Experimentos

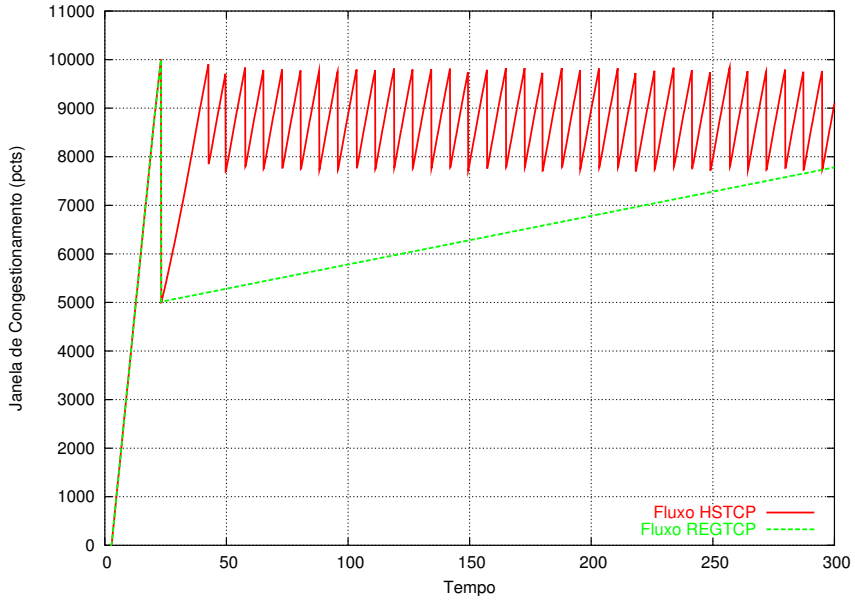
Este capítulo apresenta os resultados dos diversos experimentos realizados para atingir os objetivos propostos neste trabalho, bem como as observações mais relevantes. Neste capítulo não está contida a discussão sobre a razão da obtenção dos resultados, sendo esta avaliação deixada para ser desenvolvida no capítulo posterior.

5.1 Fluxos Isolados

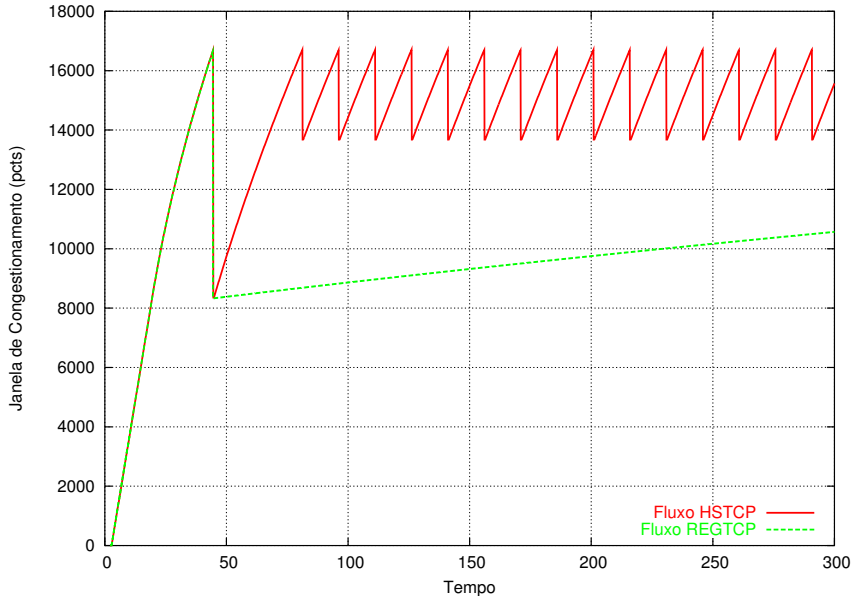
Este primeiro experimento pretendeu observar o comportamento básico de fluxos isolados REGTCP e HSTCP. Nós usamos apenas um fluxo de cada tipo, por trezentos segundos e ambos os tipos de gerenciamento de enfileiramento no roteador. O experimento foi feito apenas uma vez, sem interferência externa.

É possível observar na Figura 5.1(a) que o fluxo REGTCP possui um crescimento mais lento da janela de congestionamento que o fluxo HSTCP. O fluxo REGTCP atinge o limite da banda de 8333 pacotes em torno de 300 segundos. Por sua vez, o fluxo HSTCP atinge este ponto antes de 50 segundos.

A segunda observação importante está relacionada com a influência do tipo de gerenciamento de enfileiramento do roteador. O tamanho da janela de congestionamento atingida em ambos os tipos de TCP é maior quando DT é usado do que quando RED é usado.



(a) RED



(b) DT

Figura 5.1: Evolução da Janela de Congestionamento de um Único Fluxo

5.2 Condição Ideal

O conjunto de simulações deste experimento objetivou obter um padrão de comparação do comportamento dos fluxos REGTCP e HSTCP, quando não houvesse interferência externa, exceto o ruído de fundo (vide seção 4.3.2). Utilizamos três conjuntos de fluxos para desenvolver este experimento. O primeiro continha 1, 2, 6, 10, 20, 30 e 40 fluxos HSTCP, o segundo continha 1, 2, 6, 10, 20, 30 e 40 fluxos REGTCP e o terceiro era formado por uma mescla de fluxos REGTCP e HSTCP. Ele continha 2, 6, 10, 20, 30 e 40 fluxos, metade de cada tipo. Cada conjunto de fluxos rodou com as políticas de enfileiramento RED e DT nos roteadores. Cada simulação rodou por trezentos segundos e cada uma foi repetida dez vezes. A linha cruzando os pontos representa a mediana destas dez simulações. Cada repetição diferiu da outra apenas pelo número aleatório gerado no início de cada simulação, usado para inicialixar o gerador de números randomicos. Os resultados apresentam o desempenho das métricas selecionadas para este experimento.

A Figura 5.2 apresenta a utilização do enlace, para o primeiro e o segundo conjunto de fluxos quando o gerenciamento de enfileiramento no roteador RED foi usado. DT obteve resultados semelhantes.

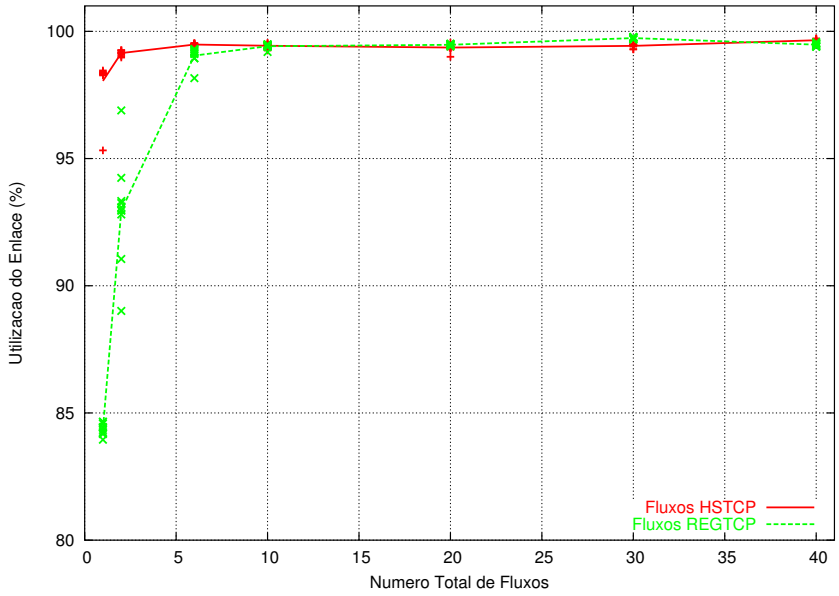


Figura 5.2: Utilização do Enlace - Condição Ideal - Fluxos Homogêneos - RED

Este gráfico mostra que os fluxos HSTCP podem atingir uma maior utilização do enlace com um número reduzido de fluxos. Muito embora os fluxos REGTCP estejam usando toda a banda disponível, fica claro que eles precisam de um número maior de fluxos para se aproximar de 100% de utilização do enlace.

Os gráficos seguintes na Figura 5.3 apresentam a taxa de eventos de congestionamento para o primeiro e segundo conjunto de fluxos, quando RED e DT são empregados, respectivamente.

Estes gráficos mostram que existe uma clara diferença entre a taxa de eventos de congestionamento resultante da utilização de cada tipo de fluxo. O uso do HSTCP produz uma taxa de eventos de congestionamento mais elevada. Outro importante aspecto a observar é que a taxa de eventos de congestionamento do HSTCP nunca é inferior a 10^{-6}.

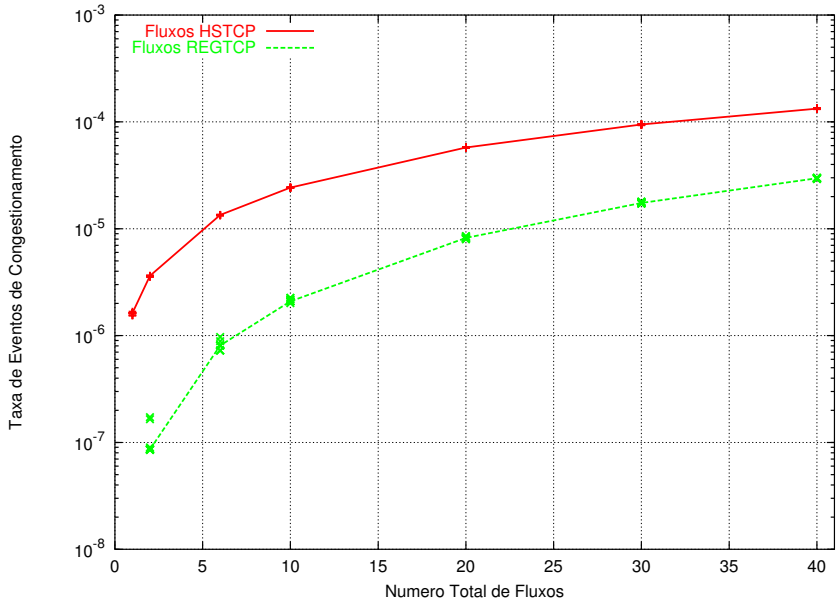
A utilização do enlace alcançada pelo terceiro conjunto de fluxos é apresentada nos gráficos da Figura 5.4. O desempenho é apresentado separadamente para cada tipo de fluxo. Uma linha é o resultado agregado de todos os fluxos HSTCP e a outra linha é o resultado agregado para os fluxos REGTCP. A terceira linha é o resultado de todos os fluxos combinados. Um gráfico mostra o desempenho quando o gerenciamento de enfileiramento do roteador RED é usado e o outro quando DT é empregado.

Estes gráficos mostram que, quando os fluxos HSTCP estão diretamente competindo com os fluxos REGTCP, a porção de banda usada pelo HSTCP é maior que a banda usada pelo fluxos REGTCP. Este fato é independente do tipo de gerenciamento de enfileiramento do roteador usado. Por outro lado, a porção de banda usada pelo HSTCP decresce quando o número total de fluxos aumenta.

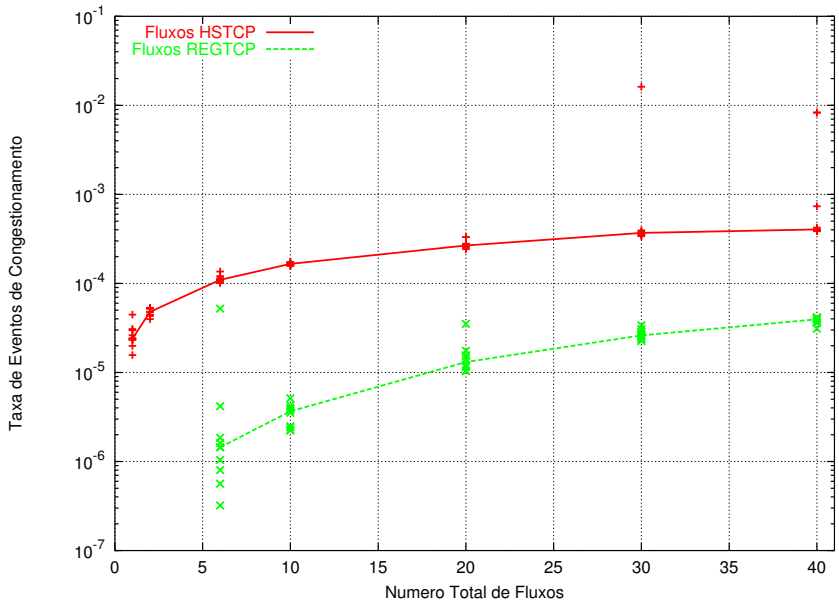
A Figura 5.5 ilustra a taxa de eventos de congestionamento observada para o terceiro conjunto de fluxos, quando os dois tipos de fluxos são empregados conjuntamente. Esta figura mostra o resultado do gerenciamento de enfileiramento do roteador RED e DT.

Este gráfico revela a evolução da taxa de eventos de congestionamento para esta situação e mostra o aumento desta taxa quando o número de fluxos aumenta.

A imparcialidade relativa para o terceiro conjunto de fluxos está retratada na Figura 5.6. Ela mostra a razão entre a quantidade de banda usada por todos os fluxos

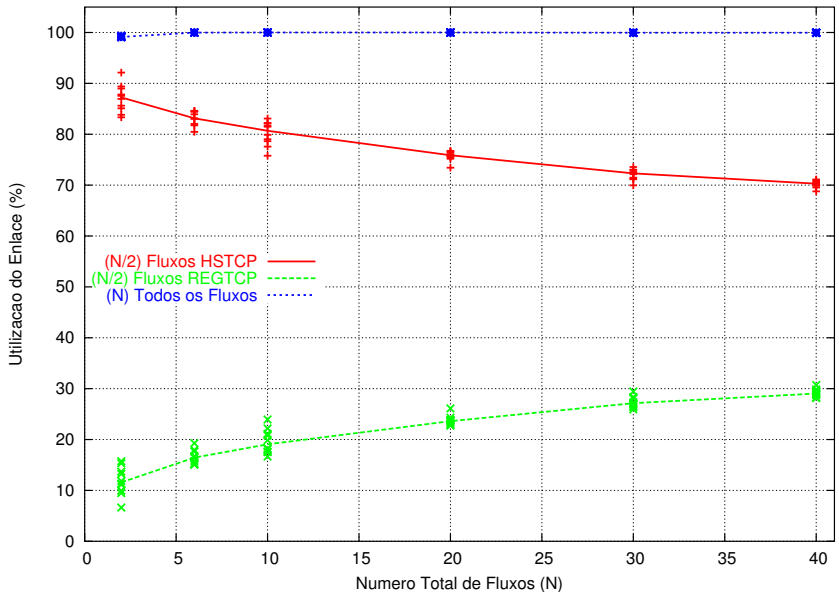


(a) RED

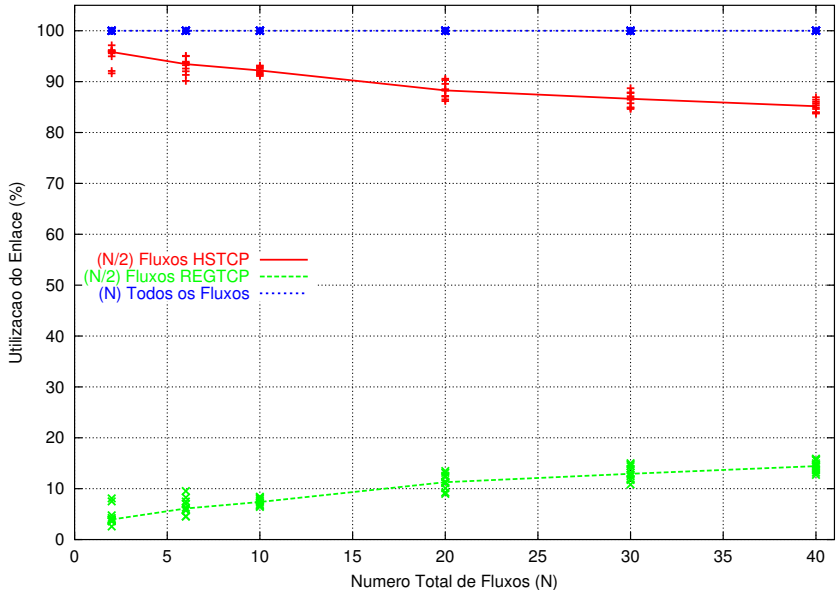


(b) DT

Figura 5.3: Taxa de Eventos de Congestionamento - Condição Ideal - Fluxos Homogêneos



(a) RED



(b) DT

Figura 5.4: Utilização do Enlace - Condição Ideal - Fluxos Heterogêneos

HSTCP e a quantidade de banda usada por todos os fluxos restantes REGTCP. Este gráfico revela que a desproporção entre a utilização do enlace entre ambos os tipos de

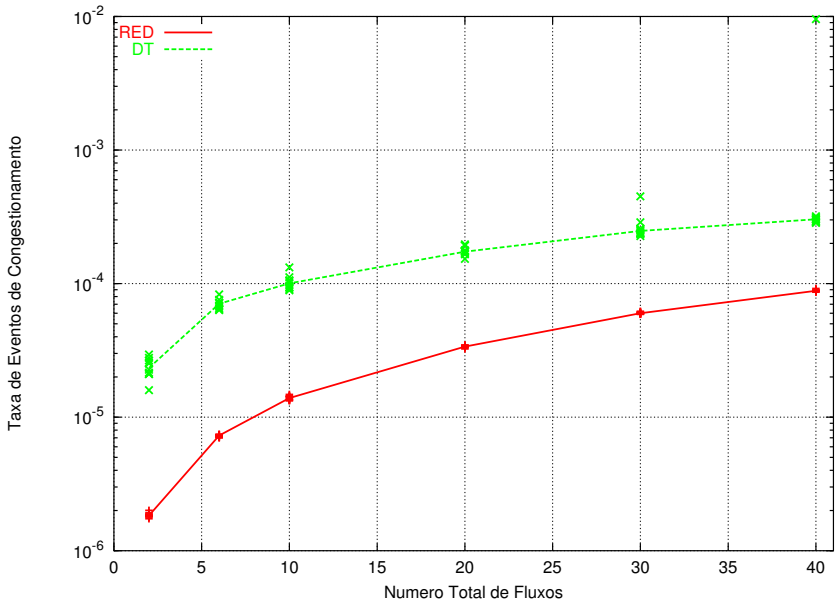


Figura 5.5: Taxa de Eventos de Congestionamento - Condição Ideal - Fluxos Heterogêneos

fluxo decresce quando o número de fluxos aumenta. Outra observação que precisa ser apontada aqui é a existência de uma ampla faixa de valores de imparcialidade quando existe um pequeno número de fluxos competindo pelo enlace. A razão encontrada neste experimento alcançou valores maiores até que 35 vezes. Também é importante observar que quando RED é empregado, a imparcialidade relativa é menor do que quando DT é usado.

O último resultado, apresentado na Figura 5.7, é a quantidade de banda roubada dos fluxos REGTCP quando eles são empregados conjuntamente com fluxos HSTCP. Este resultado é calculado usando a diferença entre a utilização do enlace alcançada por N fluxos REGTCP quando eles estão competindo contra M outros fluxos REGTCP e a utilização do enlace atingida pelo mesmo número de fluxos REGTCP quando eles estão competindo contra M outros fluxos HSTCP.

Este gráfico mostra que a quantidade de banda roubada decresce quando o número de fluxos aumenta. Este fato ressalta que a agressividade do HSTCP adapta-se à mudança nas condições de tráfego. Outra informação retratada por este gráfico é que, embora a banda roubada diminua quando o número de fluxos aumenta, a distância entre as quantidades de bandas roubadas, quando RED é usado e quando DT é usado, aumenta levemente.

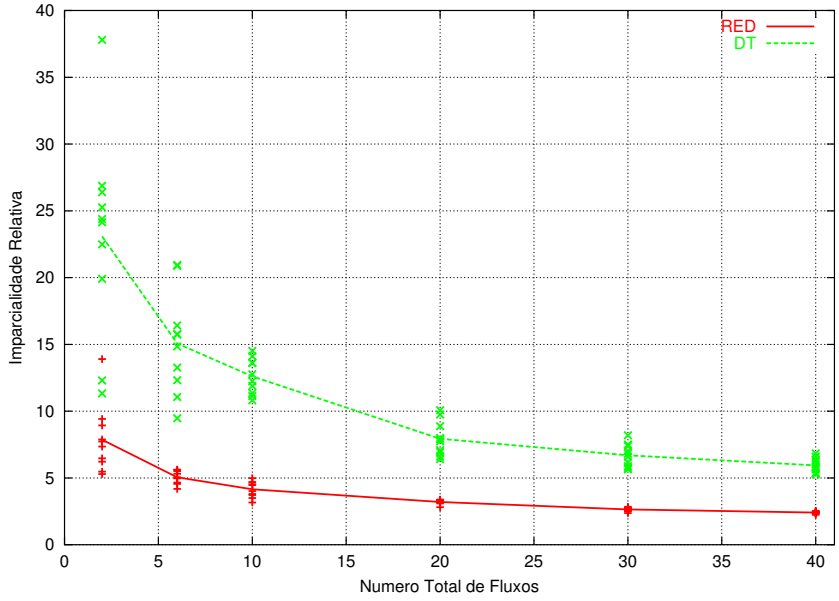


Figura 5.6: Imparcialidade Relativa - Condição Ideal - Fluxos Heterogêneos

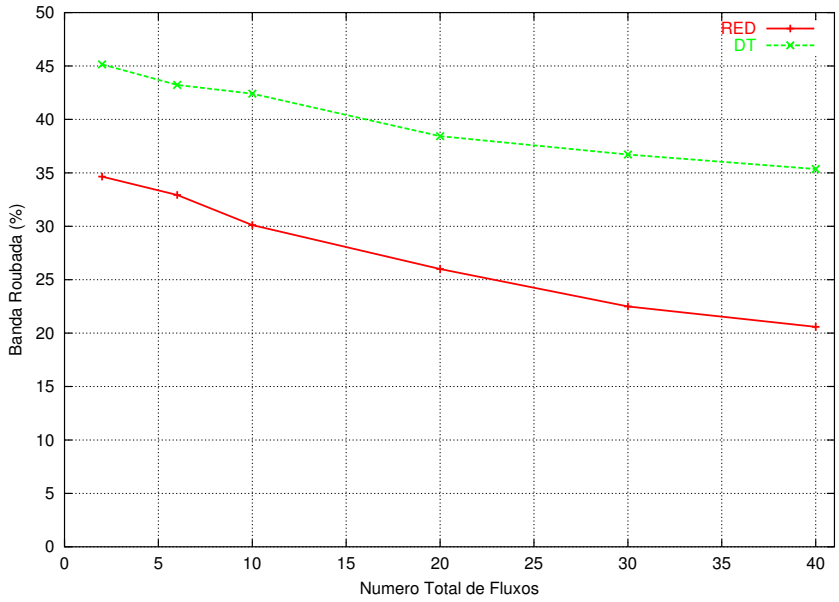


Figura 5.7: Banda Roubada - Condição Ideal

5.3 Condição de Enlace com Perdas

O foco deste conjunto de simulações foi observar o comportamento dos fluxos HSTCP e REGTCP quando sujeitos a perdas sistêmicas (perdas não devidas a conges-

tionamento). Nós usamos o modelo de erros do simulador para simular perdas no enlace gargalo, conforme explicado na seção 4.2. Este modelo de perdas foi configurado para descartar pacotes com uma taxa de descarte média definida. As taxas de perdas usadas foram 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3} e 10^{-2}. Nós utilizamos três conjuntos de fluxos para desenvolver este experimento. O primeiro continha 10 fluxos HSTCP, o segundo continha 10 fluxos REGTCP e o terceiro era formado por uma mescla de 5 fluxos REGTCP e 5 fluxos HSTCP. Cada conjunto de fluxos rodou com as políticas de enfileiramento RED e DT nos roteadores. Cada simulação rodou por trezentos segundos e cada uma foi repetida dez vezes. A linha cruzando os pontos nas figuras subsequentes representa a mediana destas dez simulações. Cada repetição diferiu da outra apenas pelo número aleatório gerado no início de cada simulação. Os resultados apresentam o desempenho das métricas selecionadas para este experimento.

A Figura 5.8 representa o desempenho da métrica utilização do enlace, para o primeiro e segundo conjuntos de fluxos quando o gerenciamento de enfileiramento de roteador RED foi usado. O DT apresentou resultados semelhantes, porque, como o enlace quase nunca estava completamente utilizado, o gerenciamento de enfileiramento do roteador quase nunca foi usado.

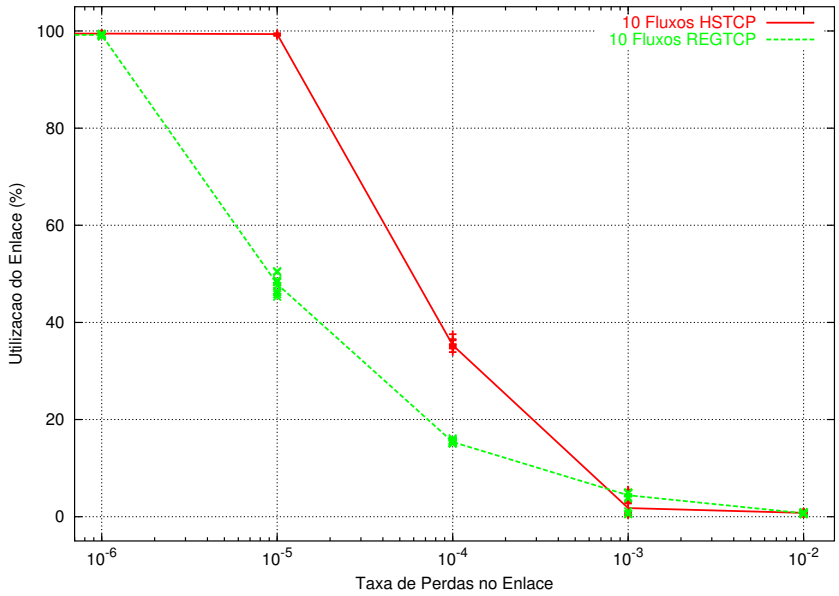


Figura 5.8: Utilização do Enlace - Condição de Enlace com Perdas - Fluxos Homogêneos - RED

O conjunto de fluxos contendo apenas REGTCP apresenta uma forte perda de desempenho quando a taxa de perdas no enlace aumenta. Este fato indica que os fluxos REGTCP não estão fazendo um uso razoável da banda disponível. Ao contrário, os fluxos HSTCP apresentam um desempenho melhor e consistentemente usam mais banda do que os fluxos REGTCP.

Os gráficos na Figura 5.9 apresentam a taxa de eventos de congestionamento para o primeiro e segundo conjunto de fluxos, quando RED e DT são empregados, respectivamente.

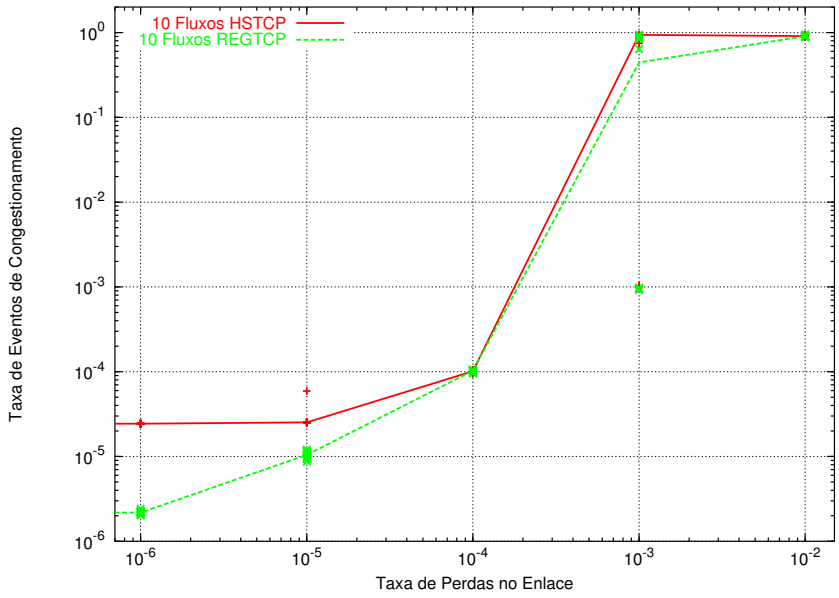
Existem dois pontos de interesse nestes gráficos. O primeiro é que a taxa de eventos de congestionamento para o conjunto de fluxos HSTCP não é menor que um limite, 10^{-5} quando RED é usado e não passa muito de 10^{-4} quando DT é empregado. Em torno deste valor, existe um ponto de inflexão na taxa de eventos de congestionamento. O segundo ponto a ressaltar é que o número de eventos de congestionamento pode aumentar para o conjunto de fluxos HSTCP perto do limite de banda, como é o caso quando DT é usado.

O grande aumento na taxa de eventos de congestionamento para uma taxa de perdas no enlace maior que 10^{-3} é devido ao grande número de retransmissões necessárias, causada pelo grande número de pacote perdidos.

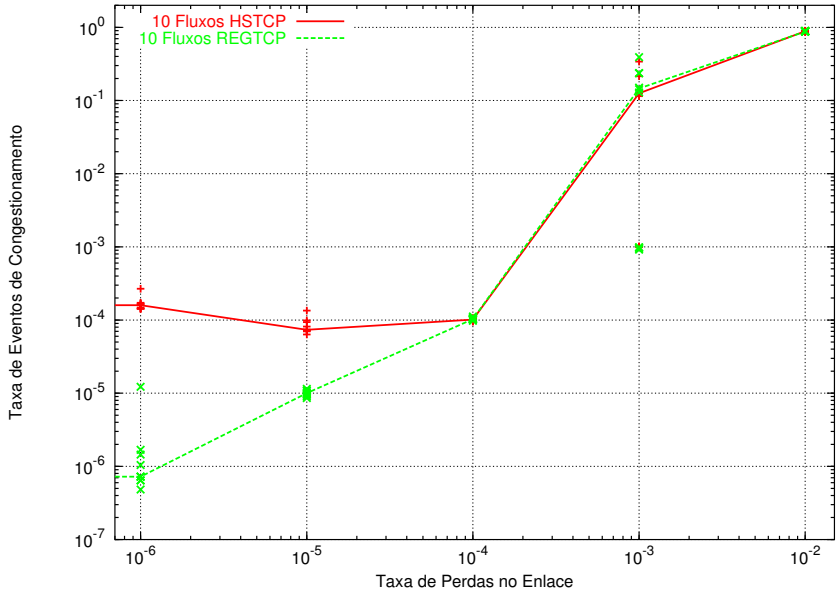
A utilização do enlace atingida pelo terceiro conjunto de fluxos é apresentada na Figura 5.10. O desempenho é apresentado separadamente para cada tipo de fluxo. Uma linha é o resultado agregado de todos os fluxos HSTCP e a outra linha é o resultado agregado para os fluxos REGTCP. A terceira linha é o resultado de todos os fluxos combinados.

Nós vemos que a diferença entre a banda que os fluxos HSTCP usam e a banda que os fluxos REGTCP são capazes de usar diminui com o aumento do número de perdas. Outro importante aspecto a ser apontado é que, para uma taxa de perdas no enlace em torno de 10^{-5}, o enlace é completamente utilizado e abaixo desta taxa, perdas por congestionamento irão dominar.

A imparcialidade relativa para o terceiro conjunto de fluxos é retratada na Figura 5.11. Ela mostra a razão entre a quantidade de banda usada por todos os fluxos HSTCP e a quantidade de banda usada por todos os fluxos restantes REGTCP.

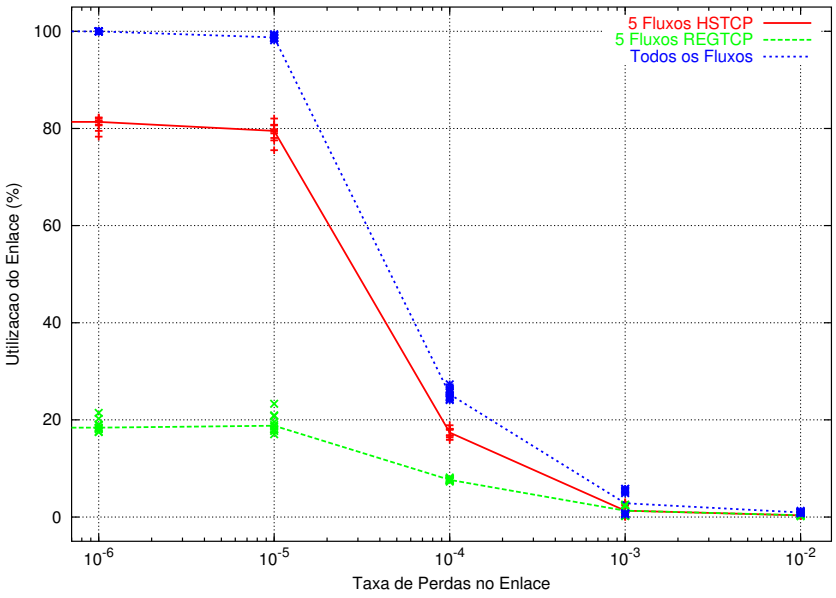


(a) RED

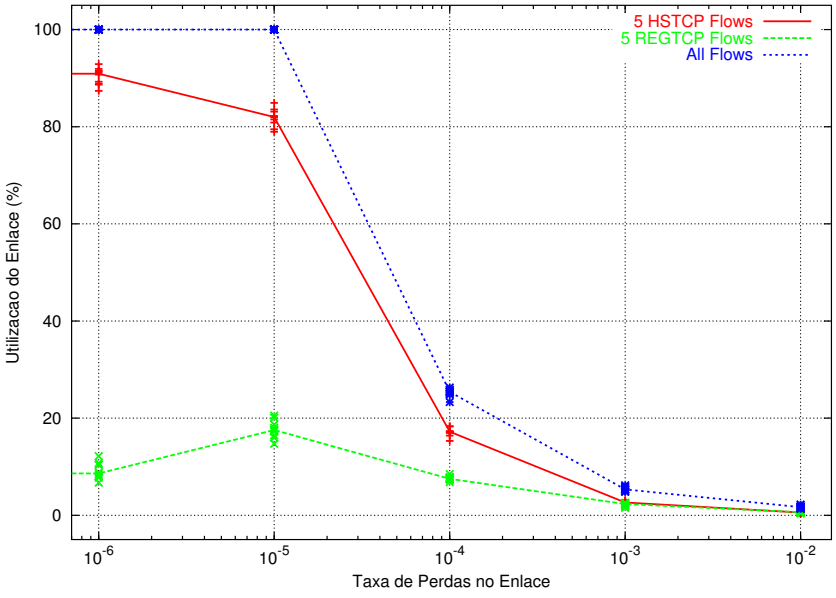


(b) DT

Figura 5.9: Taxa de Eventos de Congestionamento - Condição de Enlace com Perdas - Fluxos Homogêneos



(a) RED



(b) DT

Figura 5.10: Utilização do Enlace - Condição de Enlace com Perdas - Fluxos Heterogêneos

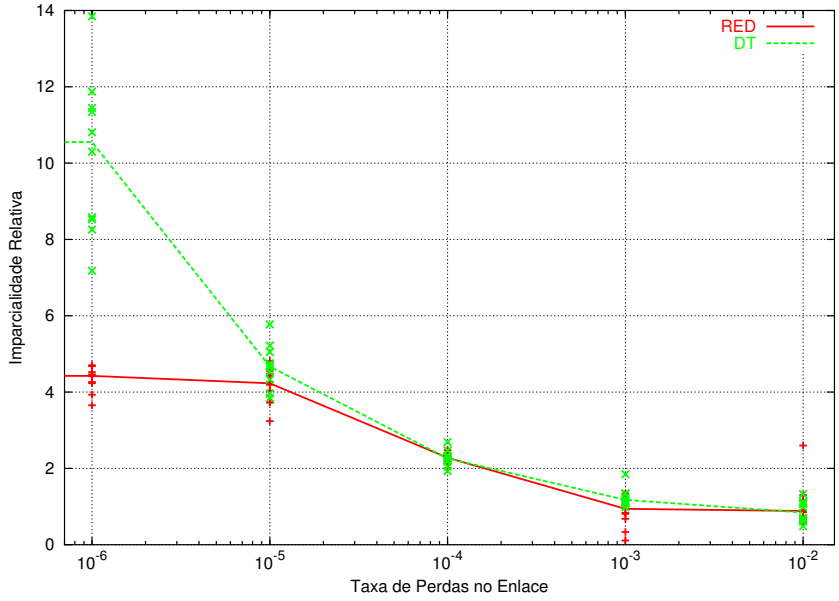


Figura 5.11: Imparcialidade Relativa - Condição de Enlace com Perdas - Fluxos Heterogêneos

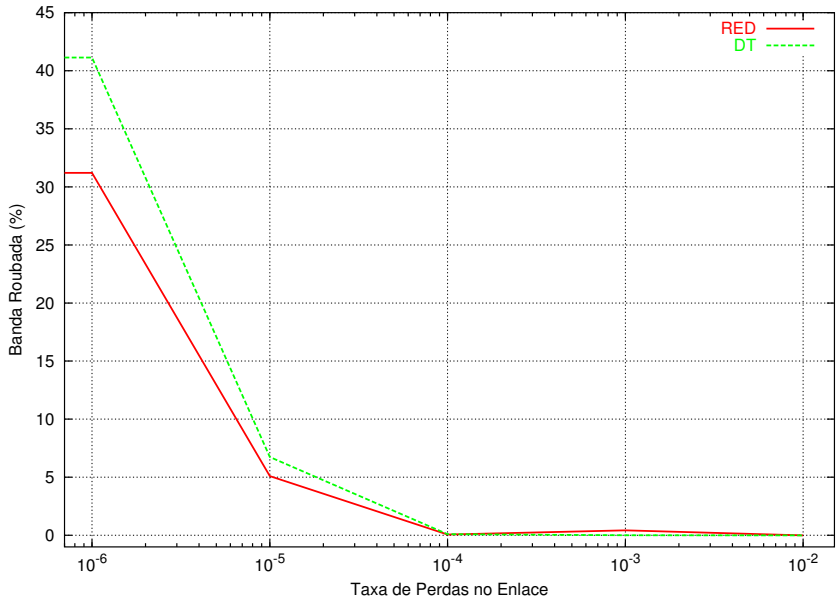


Figura 5.12: Banda Roubada - Condição de Enlace com Perdas

A informação sobre a quantidade de banda roubada de todos os fluxos REGTCP quando eles são empregados conjuntamente com os fluxos HSTCP está apresentada na Figura 5.12. Este resultado é calculado usando a diferença entre a utilização do enlace alcançada por um número de fluxos REGTCP quando eles estão competindo contra M

outros fluxos REGTCP e a utilização do enlace atingida pelo mesmo número de fluxos REGTCP quando eles estão competindo contra M outros fluxos HSTCP.

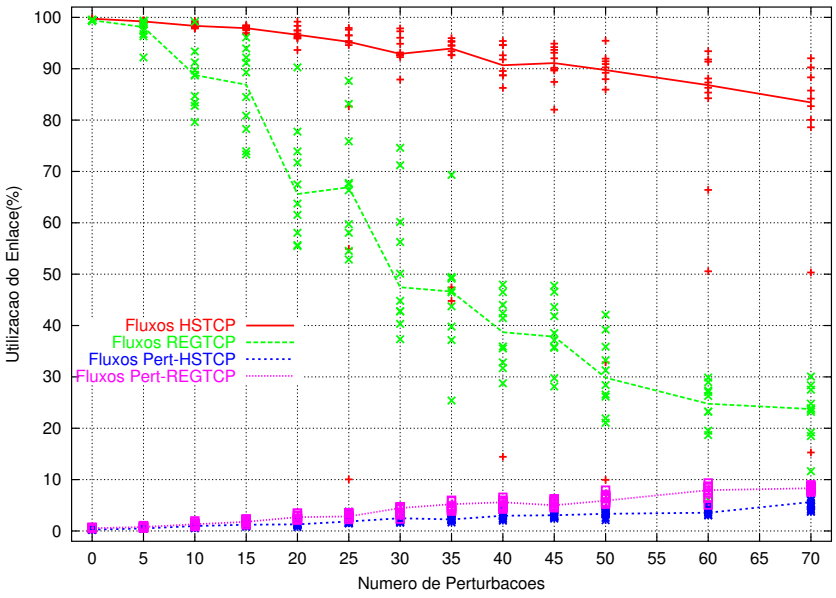
Este gráfico mostra que os fluxos REGTCP não perdem banda devido ao emprego dos fluxos HSTCP quando a taxa de perdas no enlace é maior que 10^{-4}. Para taxas menores que este nível, a quantidade banda roubada é perceptível.

5.4 Condição de Tráfego em Rajada

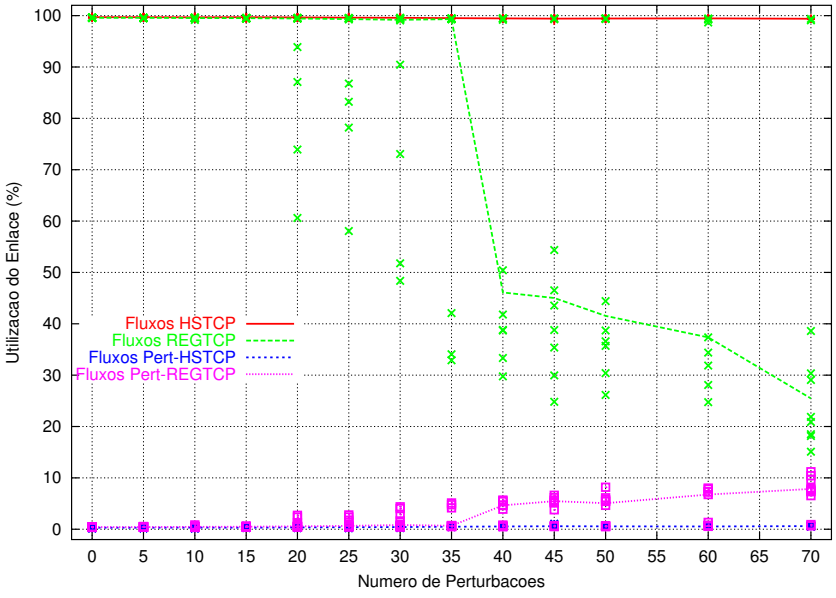
A meta deste conjunto de experimentos foi de entender o comportamento e reação dos fluxos REGTCP e dos fluxos HSTCP, quando eles são submetidos a tráfego em rajada. O tráfego em rajada foi composto por fluxos de TCP Padrão de curta duração que duram apenas poucos segundos, de forma a só rodarem durante a fase de Partida Lenta. Como a fase de Partida Lenta possui um crescimento exponencial, os fluxos com característica de rajada têm um considerável impacto nos fluxos de longa duração. Nós usamos 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60 e 70 fluxos em rajada distribuídos randomicamente, com distribuição uniforme, durante todo o período de simulação.

Nós utilizamos três conjuntos de fluxos para desenvolver este experimento. O primeiro continha 10 fluxos HSTCP, o segundo continha 10 fluxos REGTCP e o terceiro era formado por uma mescla de 5 fluxos REGTCP e 5 fluxos HSTCP. Cada conjunto de fluxos rodou com as políticas de enfileiramento RED e DT nos roteadores. Cada simulação rodou por trezentos segundos e cada uma foi repetida dez vezes. A linha cruzando os pontos representa a mediana destas dez simulações. Cada repetição diferiu da outra apenas pelo número aleatório gerado no início de cada simulação. Os resultados apresentam o desempenho das métricas selecionadas para este experimento.

Os gráficos na Figura 5.13 apresentam o desempenho da métrica de utilização do enlace agregada, para o primeiro e segundo conjunto de fluxos quando os gerenciamentos de enfileiramento do roteador RED e DT foram usados. Estes gráficos também apresentam a utilização agregada do enlace para os fluxos em rajada (ou perturbações) presentes quando o primeiro conjunto de fluxos foi usado e também quando o conjunto de fluxos REGTCP rodou.



(a) RED



(b) DT

Figura 5.13: Utilização do Enlace - Condição de Tráfego em Rajada - Fluxos Homogêneos

Nós observamos na Figura 5.13(a) que o conjunto de fluxos HSTCP decresce sua utilização do enlace suavemente enquanto o número de perturbações aumenta. Por outro lado, o impacto no conjunto de fluxos REGTCP é maior e seu desempenho diminui rapidamente quando o número de perturbações aumenta.

Outra informação fornecida pelo primeiro gráfico é que a quantidade de banda utilizada pelas pertubações quando competindo contra fluxos HSTCP é ligeiramente inferior à banda usada quando competindo contra o conjunto de fluxos REGTCP.

O impacto do uso de um gerenciamento de enfileiramento de roteador distinto fica claro quando o conjunto de fluxos HSTCP é submetido a um tráfego em rajadas. A utilização do enlace decresce levemente com RED, mas os fluxos HSTCP são quase imunes à perturbação quando a política de enfileiramento de roteador DT é usada, como pode ser visto na Figura 5.13(b).

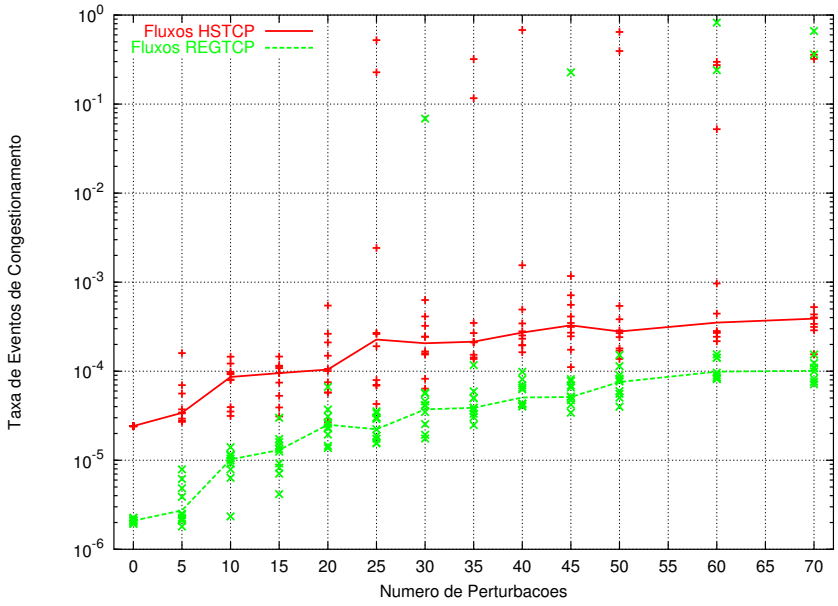
Os próximos dois gráficos na Figura 5.14 apresentam a taxa de eventos de congestionamento agregada para o primeiro e segundo conjunto de fluxos, quando RED e DT foram empregados.

Nós observamos que a taxa de eventos de congestionamento aumenta continuamente quando o número de perturbações aumenta, durante o uso do RED como gerenciamento de enfileiramento do roteador. Este comportamento acontece com o conjunto de fluxos HSTCP bem como com o conjunto de fluxos REGTCP. Quando o gerenciamento de enfileiramento do roteador DT é empregado, o comportamento muda. O conjunto de fluxos HSTCP apresenta uma taxa de eventos de congestionamento quase constante e o conjunto de fluxos REGTCP apresenta dois níveis para taxa de eventos de congestionamento, provavelmente causado por sincronização global.

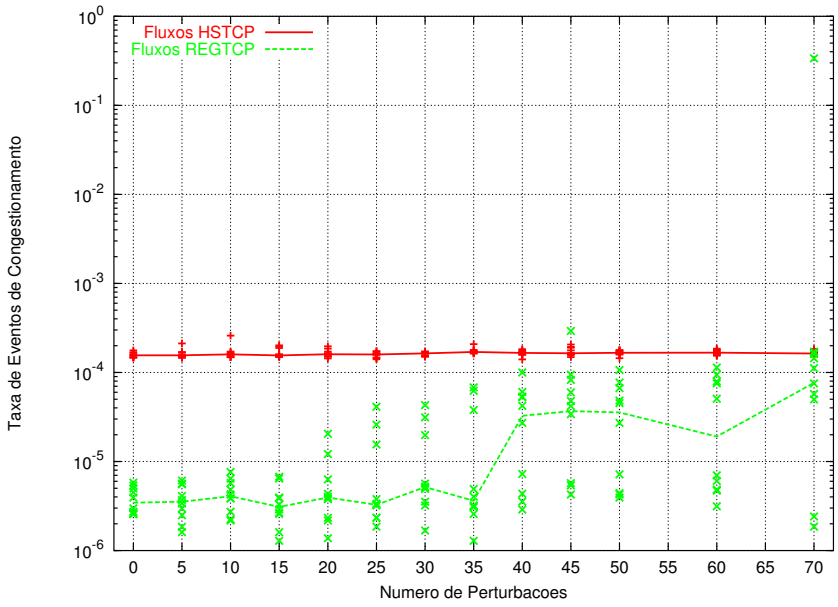
A Figura 5.15 apresenta o número absoluto de eventos de congestionamento (pacotes descartados mais os pacotes com ECN marcado) para o caso quando o gerenciamento de enfileiramento do roteador RED é usado.

Este gráfico revela que o número de eventos de congestionamento acontecidos, quando o conjunto de fluxos REGTCP é usado, foi inferior ao número de quando o conjunto de fluxos HSTCP foi empregado.

A utilização do enlace alcançada pelo terceiro conjunto de fluxos está apresentada na Figura 5.16. Ela mostra o desempenho separadamente para cada tipo de fluxo.



(a) RED



(b) DT

Figura 5.14: Taxa de Eventos de Congestionamento - Condição de Tráfego em Rajada - Fluxos Homogêneos

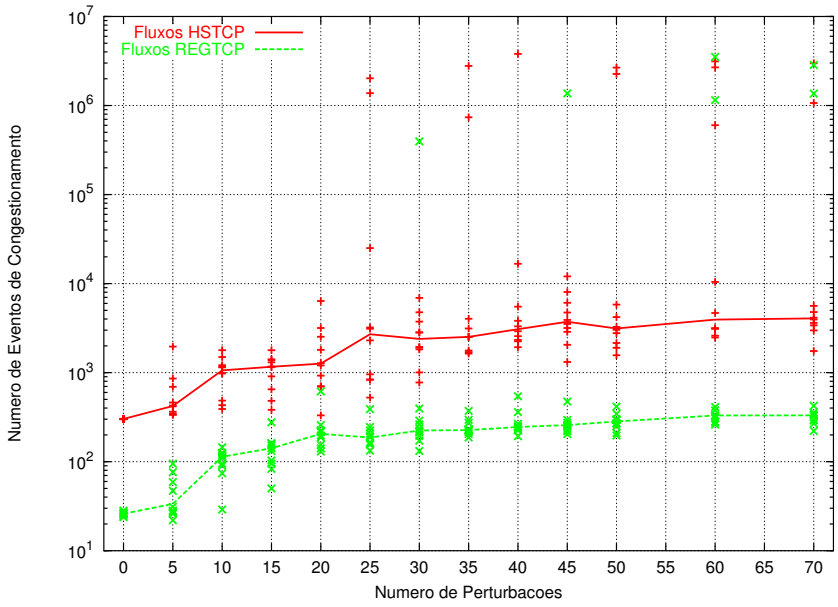


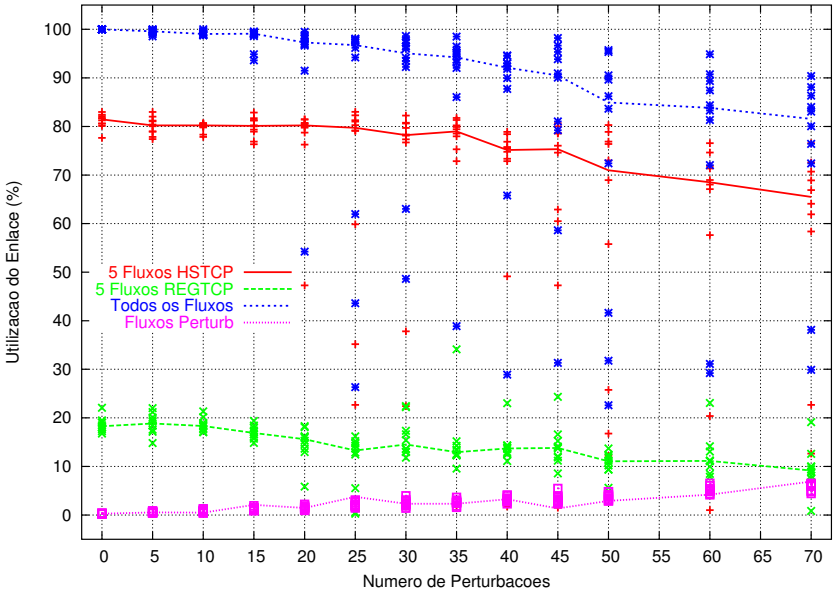
Figura 5.15: Eventos de Congestionamento - Condição de Tráfego em Rajadas - Fluxos Homogêneos - RED

Uma linha é o resultado agregado do conjunto de fluxos HSTCP e a outra linha é o resultado agregado para o conjunto de fluxos REGTCP. A terceira linha é o resultado de todos os fluxos combinados. A linha restante representa a utilização de enlace agregada para todas as perturbações. Um gráfico mostra o desempenho quando o gerenciamento de enfileiramento do roteador RED é usado e o outro quando DT é empregado.

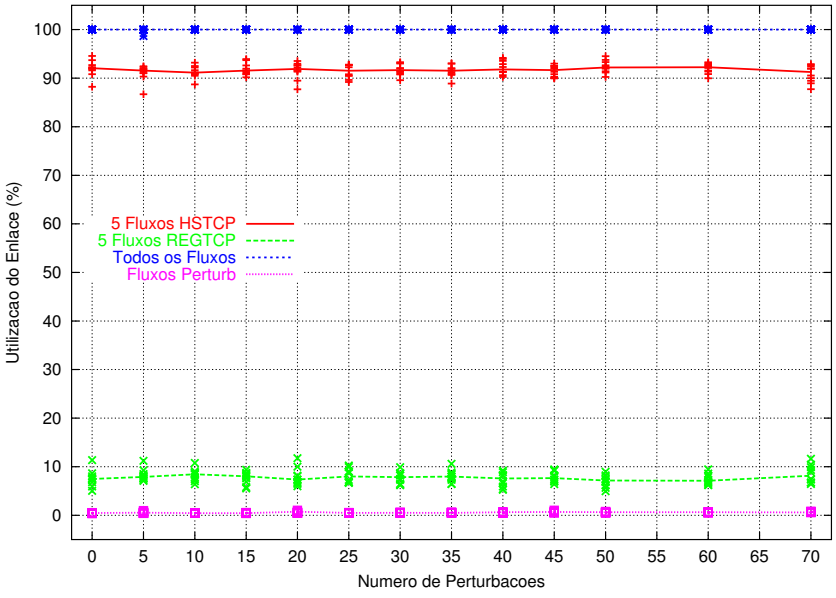
A informação importante fornecida por estes gráficos é o pobre e quase constante desempenho do conjunto de fluxos REGTCP. Ele tem uma baixa utilização do enlace, entretanto, este desempenho não muito alterado quando mais perturbações são usadas.

A imparcialidade relativa para o terceiro conjunto de fluxos é retratada na Figura 5.17. Ela mostra a razão entre a quantidade de banda usada por todos os fluxos HSTCP e a quantidade de banda usada por todos os fluxos restantes REGTCP.

A imparcialidade relativa é quase constante usando-se RED, bem como quando DT é utilizado. Os fluxos HSTCP ocupam de 10 a 15 vezes mais banda que os fluxos REGTCP, quando DT é empregado, porém quando RED é usado, este valor cai para em torno de 5 vezes mais. O nível de imparcialidade relativa com RED cresce levemente quando o número de perturbações aumenta.



(a) RED



(b) DT

Figura 5.16: Utilização de Enlace - Condição de Tráfego em Rajadas - Fluxos Heterogêneos

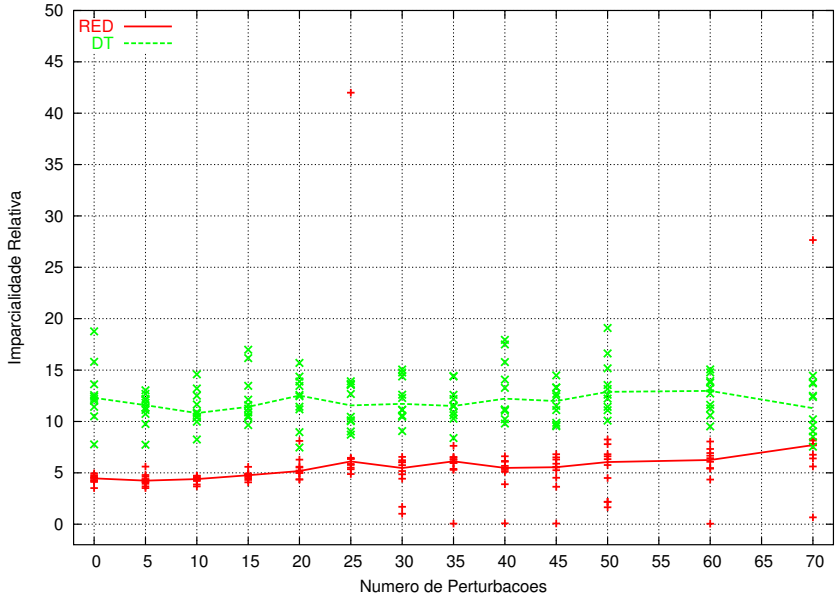


Figura 5.17: Imparcialidade Relativa - Condição de Tráfego em Rajadas - Fluxos Heterogêneos

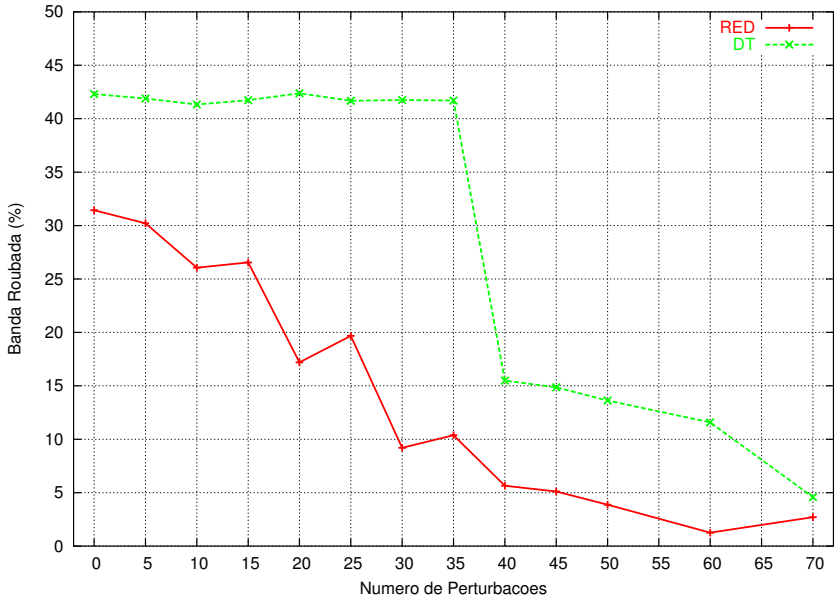


Figura 5.18: Banda Roubada - Condição de Tráfego em Rajadas

A banda roubada de todos os fluxos REGTCP quando eles são empregados em conjunto com os fluxos HSTCP está apresentada na Figura 5.18.

Esta figura ressalta que a quantidade de banda roubada pelos fluxos HSTCP dos fluxos REGTCP decresce com o aumento do número de perturbações, independente do tipo de gerenciamento de enfileiramento do roteador usado. Deve-se observar porém que a quantidade de banda roubada é maior para o gerenciamento de enfileiramento do roteador DT.

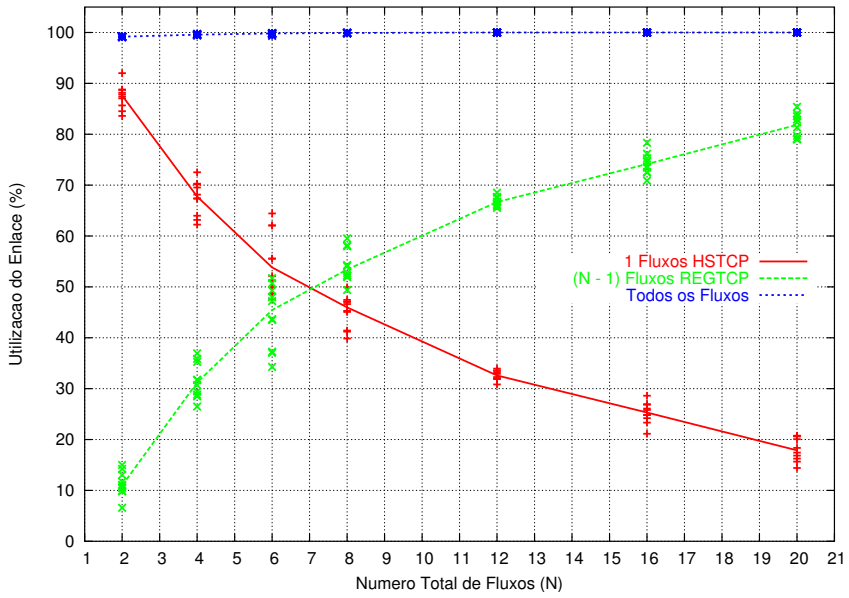
5.5 Competição Entre Fluxos Heterogêneos

No caso deste experimento, nós queríamos verificar o comportamento dos fluxos HSTCP e REGTCP quando um número assimétrico de fluxos fosse empregado. Para atingir este objetivo, nós estabelecemos executar 1 fluxo HSTCP contra um número variável de fluxos REGTCP. Nós usamos 1, 3, 5, 7, 11, 15 e 19 fluxos REGTCP. Estes fluxos rodaram com as políticas de enfileiramento RED e DT nos roteadores. Cada simulação foi executada por trezentos segundos e cada uma foi repetida dez vezes. A linha cruzando os pontos representa a mediana destas dez simulações. Cada repetição diferiu da outra apenas pelo número aleatório gerado no início de cada simulação. Os resultados apresentam o desempenho das métricas selecionadas para este experimento.

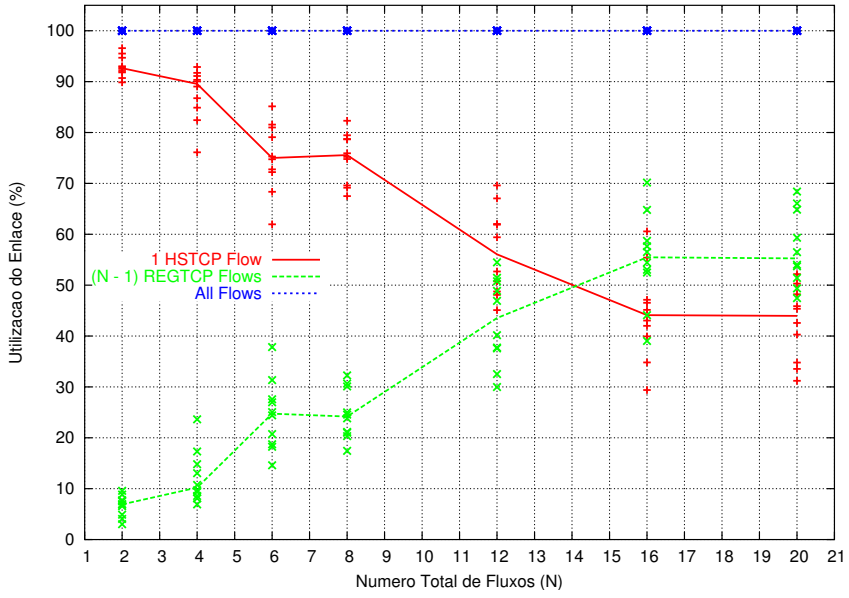
Os gráficos na Figura 5.19 apresentam o desempenho da métrica de utilização do enlace agregada quando os gerenciamentos de enfileiramento do roteador RED e DT foram usados, respectivamente. Também é apresentado a utilização do enlace agregada para todos os fluxos conjuntamente.

Algumas informações importantes estão presentes nestes gráficos. A primeira informação é que o fluxo HSTCP adapta-se com a quantidade de fluxos REGTCP utilizados e evita que o enlace fique ocioso. A segunda informação fornecida é que existe um determinado número de fluxos REGTCP que tem um desempenho equivalente a 1 fluxo HSTCP. Todavia, este número depende do tipo de gerenciamento de enfileiramento do roteador utilizado. Esta equivalência ocorre no cruzamento da linha de utilização do HSTCP com a linha de utilização dos fluxos REGTCP.

A imparcialidade relativa é apresentada na Figura 5.20. Ela mostra a razão entre a quantidade de banda usada pelo fluxo HSTCP e quantidade de banda usada por todos os fluxos REGTCP.



(a) RED



(b) DT

Figura 5.19: Utilização de Enlace - Competição Entre Fluxos Heterogêneos

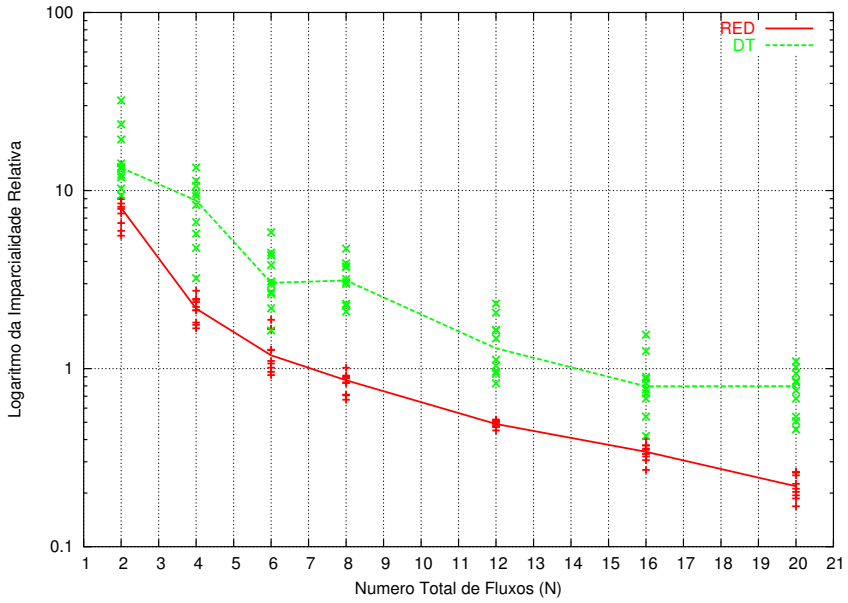


Figura 5.20: Imparcialidade Relativa - Competição Entre Fluxos Heterogêneos

5.6 Enlace com Perda Constante de 10^{-5}

Este conjunto de experimentos repete o experimento da Condição Ideal, exceto que ele introduz uma perda constante no enlace de 10^{-5}. O propósito desta mudança foi investigar o comportamento dos fluxos HSTCP e REGTCP com perdas sistêmicas. Diferentemente do experimento com a Condição de Enlace com Perdas, no qual cada conjunto de fluxos possuíam apenas 10 fluxos, nós usamos um número variável para conjunto de fluxos.

Nós utilizamos três conjuntos de fluxos para desenvolver este experimento. O primeiro continha 1, 2, 6, 10, 20, 30 e 40 fluxos HSTCP, o segundo continha 1, 2, 6, 10, 20, 30 e 40 fluxos REGTCP e o terceiro era formado por uma mescla de fluxos REGTCP e HSTCP. Ele continha 2, 6, 10, 20, 30 e 40 fluxos, metade de cada tipo. Cada conjunto de fluxos rodou com as políticas de enfileiramento RED e DT nos roteadores. Cada simulação foi executada por trezentos segundos e cada uma foi repetida dez vezes. A linha cruzando os pontos representa a mediana destas dez simulações. Cada repetição diferiu da outra apenas pelo número aleatório gerado no início de cada simulação. Os resultados apresentam o desempenho das métricas selecionadas para este experimento.

A Figura 5.21 apresenta o desempenho da métrica de utilização do enlace agregada para o primeiro e segundo conjunto de fluxos quando o gerenciamento de enfileiramento do roteador RED foi usado. DT apresenta resultados similares.

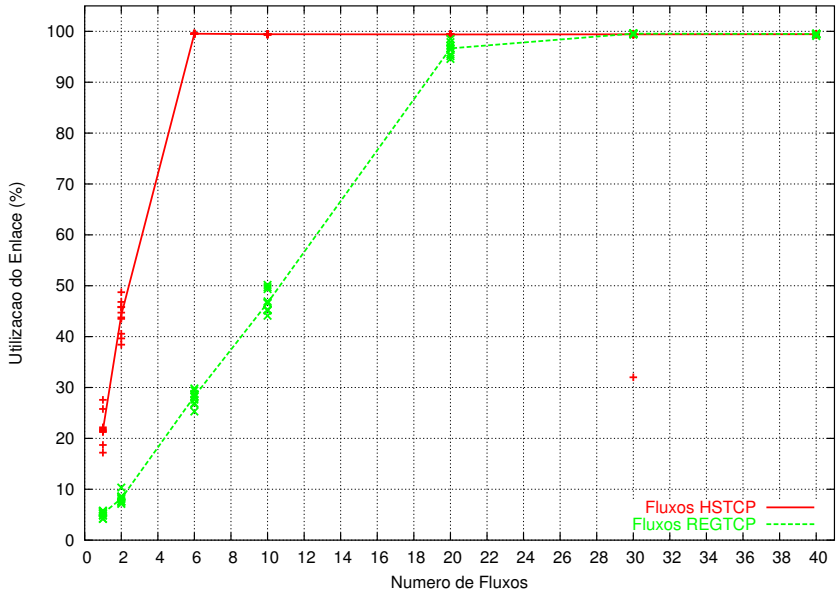


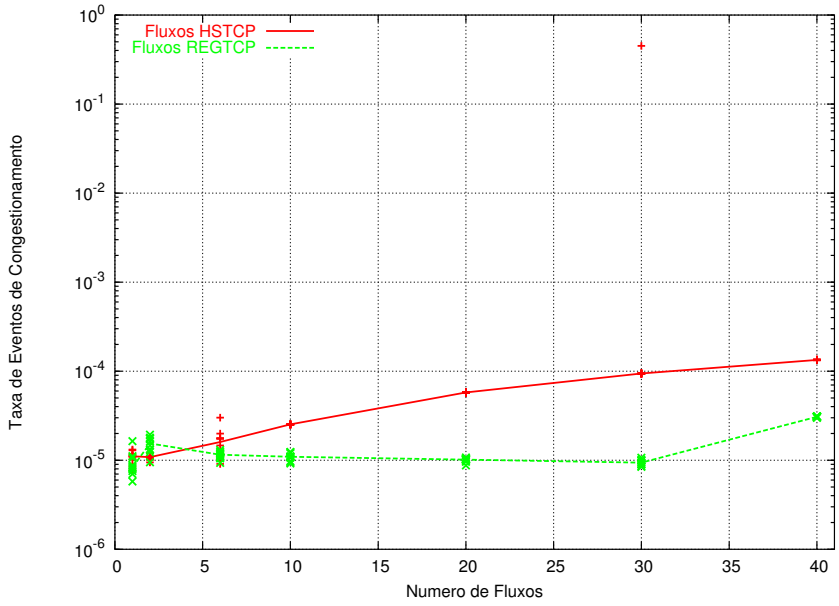
Figura 5.21: Utilização do Enlace Agregada - Enlace com Perda Constante de 10^{-5} - Fluxos Homogêneos - RED

A informação fornecida por este gráfico é que quando os fluxos HSTCP são empregados nesta condição de rede, existe a necessidade de apenas 6 fluxos para atingir plena utilização do enlace. Todavia, quando fluxos REGTCP são usados, este número aumenta para 20 fluxos ou mais.

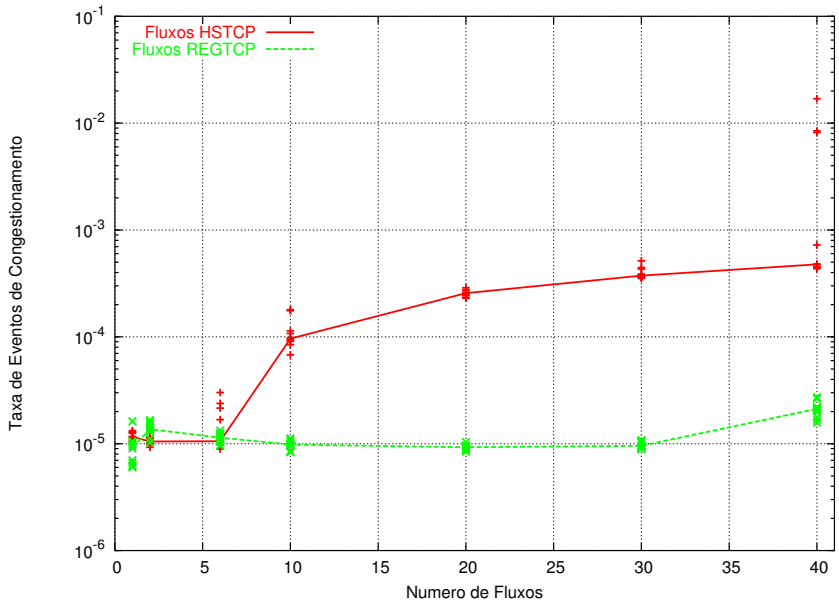
Os dois gráficos na Figura 5.22 apresentam a taxa de eventos de congestionamento para o primeiro e segundo conjunto de fluxos, quando RED e DT são empregados, respectivamente.

A taxa de eventos de congestionamento para ambos os conjuntos de fluxos apresenta uma mudança no seu comportamento. Esta mudança pode ser vista quando existem mais de 30 fluxos REGTCP e quando existem mais de 6 fluxos HSTCP, quando DT é usado. Este ponto de inflexão representa o ponto quando a plena utilização do enlace é atingida.

A utilização do enlace para o terceiro conjunto de fluxos é apresentada na Figura 5.23. Ela apresenta o desempenho separadamente para cada tipo de de fluxo. Uma

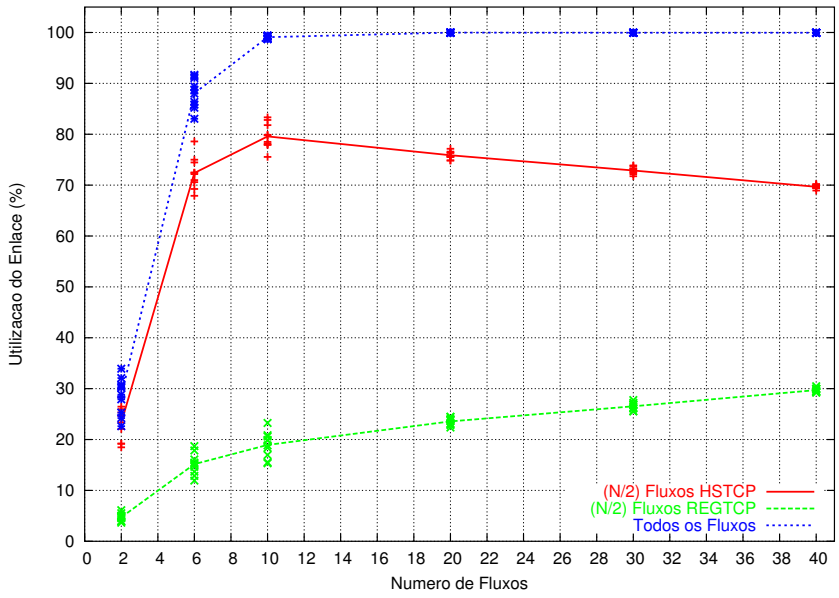


(a) RED

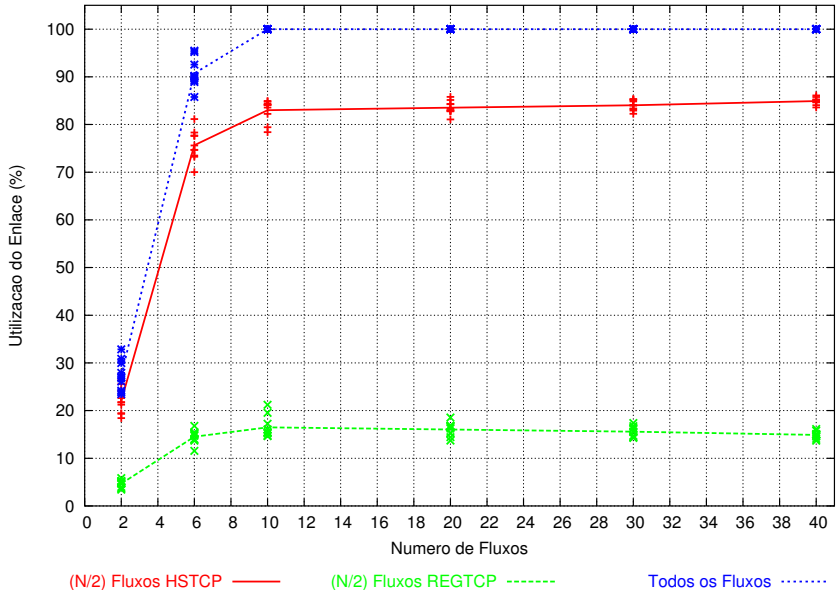


(b) DT

Figura 5.22: Taxa de Eventos de Congestionamento - Enlace com Perda Constante de 10^{-5} - Fluxos Homogêneos



(a) RED



(b) DT

Figura 5.23: Utilização do Enlace - Enlace com Perda Constante de 10^{-5} - Fluxos Heterogêneos

linha é o resultado agregado do conjunto de fluxos HSTCP e a outra linha é o resultado agregado para o conjunto de fluxos REGTCP. A terceira linha é o resultado de todos os fluxos combinados. Um gráfico mostra o desempenho quando o gerenciamento de enfileiramento do roteador RED é usado e o outro quando DT é empregado.

Estes gráficos mostram a influência do gerenciamento de enfileiramento do roteador no comportamento da utilização do enlace para cada tipo de fluxo. Enquanto no RED a utilização do enlace para os fluxos HSTCP decresce quando o número total de fluxos aumenta, no DT a mesma utilização permanece constante ou aumenta levemente.

A imparcialidade relativa para o terceiro conjunto de fluxos é mostrada na Figura 5.24. Ela apresenta a razão entre a quantidade de banda utilizada por todos os fluxos HSTCP e a quantidade de banda usada por todos os restantes fluxos REGTCP.

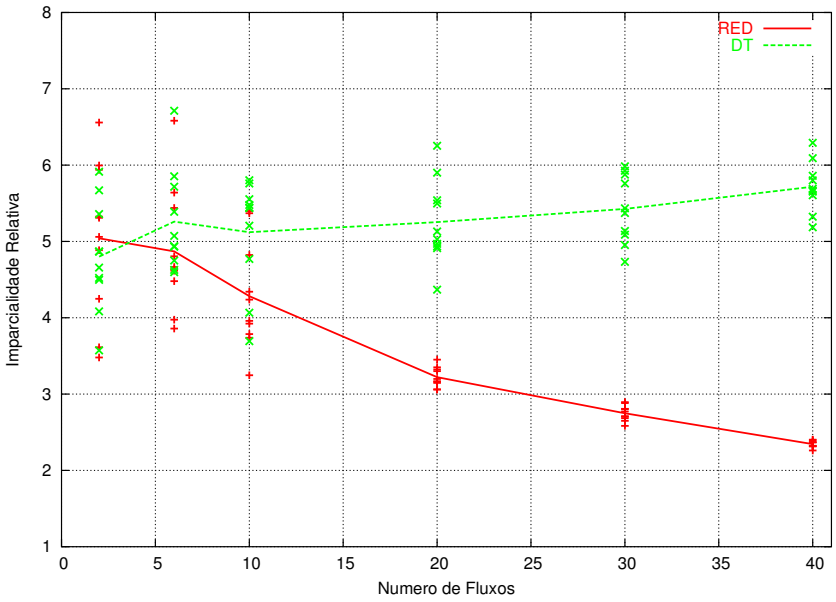


Figura 5.24: Imparcialidade Relativa - Enlace com Perda Constante de 10^{-5} - Fluxos Heterogêneos

O último resultado deste experimento é apresentado na Figura 5.25. Ele mostra a quantidade de banda roubada de todos os fluxos REGTCP quando empregados em conjunto com os fluxos HSTCP.

Das Figura 5.24 e 5.25 nós podemos ver que os fluxos HSTCP estão obtendo mais porção de banda quando o número de fluxos aumenta e quando DT é usado. O gerenciamento de enfileiramento do roteador RED apresenta resultado oposto a este.

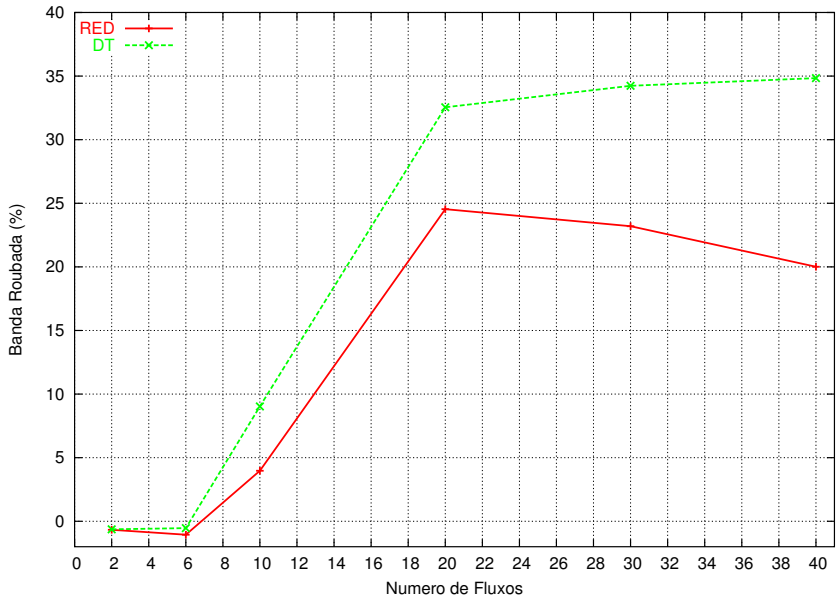


Figura 5.25: Banda Roubada - Enlace com Perda Constante de 10^{-5}

5.7 Simulação de Longa Duração

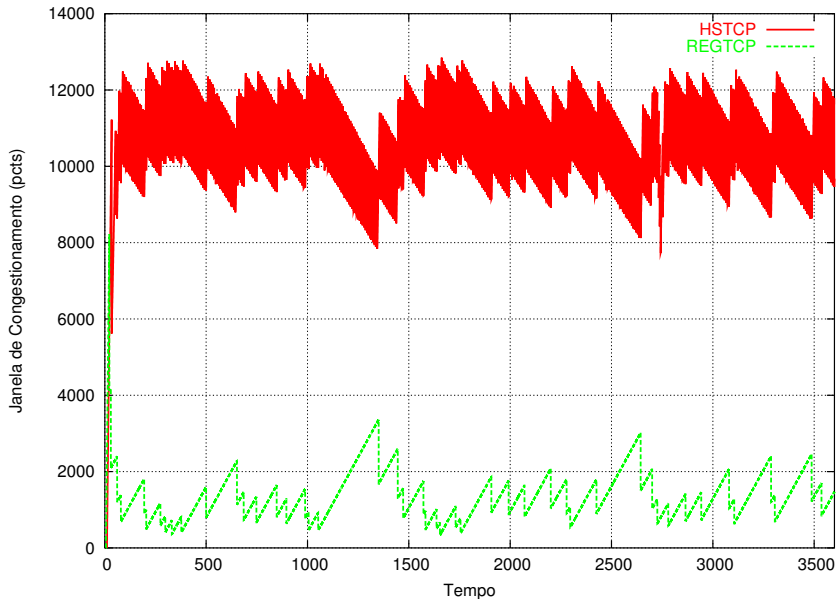


Figura 5.26: Simulação de Longa Duração - 1 Hora - RED

Este experimento ilustra a interação entre 1 fluxo HSTCP e 1 fluxo REGTCP em um longo período de tempo. Nós executamos este experimento por um período de 3600

segundos de simulação, usando o gerenciamento de enfileiramento RED. A Figura 5.26 mostra o comportamento da janela de congestionamento durante este período de tempo.

Esta figura mostra que a interação entre os dois fluxos não apresenta variação significativa em um período longo de tempo. As janelas de congestionamento permanecem em torno do mesmo nível durante todo o período. A ampla área ocupada pela linha do HSTCP representa a oscilação de sua janela de congestionamento.

5.8 Transferência por Fluxos Paralelos

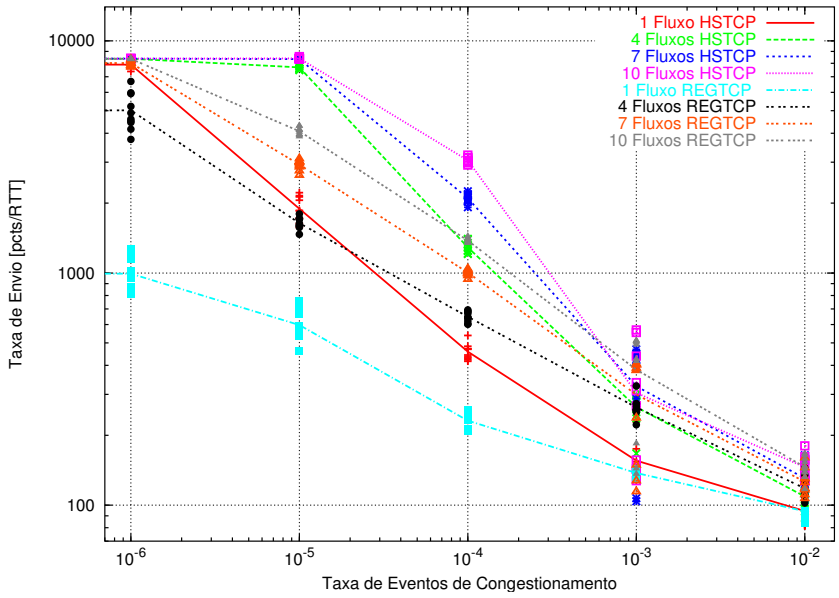


Figura 5.27: Função Resposta - Transferência por Fluxos Paralelos - RED

O foco deste conjunto de experimentos foi verificar o desempenho teórico de fluxos paralelos quando trabalhando em diversas condições de perda. Nós usamos o modelo de perdas do simulador para simular as perdas no enlace gargalo. Este modelo de perdas foi configurado para descartar pacotes com uma taxa média definida. A taxa de perdas usada foi 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3} e 10^{-2}. Nós utilizamos fluxos paralelos contendo 1, 4, 7 e 10 fluxos. O primeiro conjunto de fluxos continha apenas fluxos REGTCP e no segundo havia apenas fluxos HSTCP. Os resultados são apresentados separadamente para cada número de fluxos. Cada conjunto de fluxos foi simulado com as políticas de enfileiramento RED e DT nos roteadores. Cada simulação foi executada por trezentos

segundos e cada uma foi repetida dez vezes. A linha cruzando os pontos representa a mediana destas dez simulações. Cada repetição diferiu da outra apenas pelo número aleatório gerado no início de cada simulação.

A Figura 5.27 apresenta o desempenho de cada transferência em termos de sua taxa de envio. A unidade usada aqui foi *pacotes/RTT* e o gerenciamento de enfileiramento do roteador RED foi usado. O uso de DT apresentou resultados similares.

As Figuras 5.28 e 5.29 foram adicionadas para apresentar o desempenho teórico esperado do uso de fluxos paralelos e sua comparação com o desempenho teórico de 1 fluxo HSTCP.

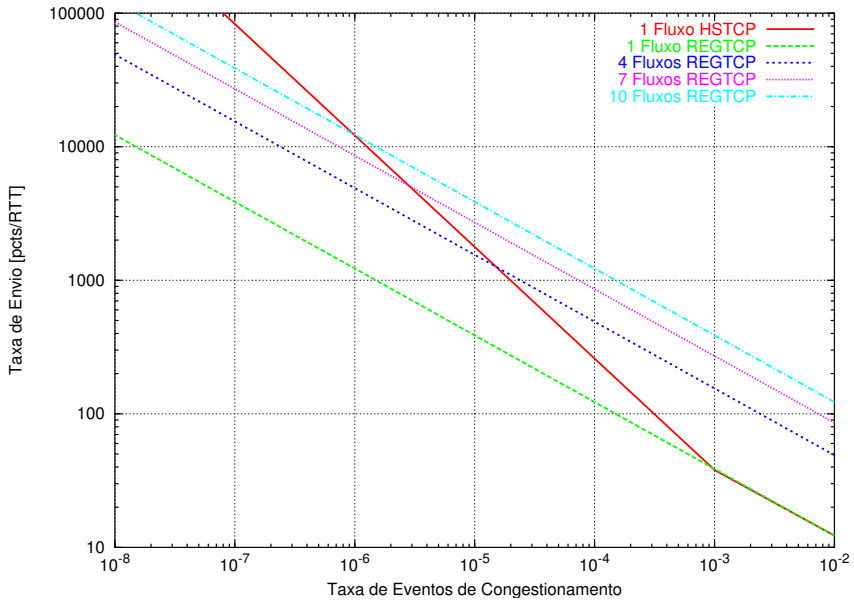


Figura 5.28: Função Resposta Teórica - Transferência por Fluxos Paralelos

O desempenho teórico de fluxos paralelos sobre uma grande faixa de taxa de eventos de congestionamento segue a equação apresentada em [25], para a condição deste experimento (MSS = 1500 bytes, RTT = 100ms, C = 1 e as perdas de pacotes impactando os fluxos paralelos na mesma extensão). Para a função resposta do HSTCP foi usada a equação definida em [18]. A Figura 5.29 apresenta a imparcialidade esperada para os fluxos paralelos e para 1 fluxo HSTCP relativamente a 1 fluxo TCP Padrão.

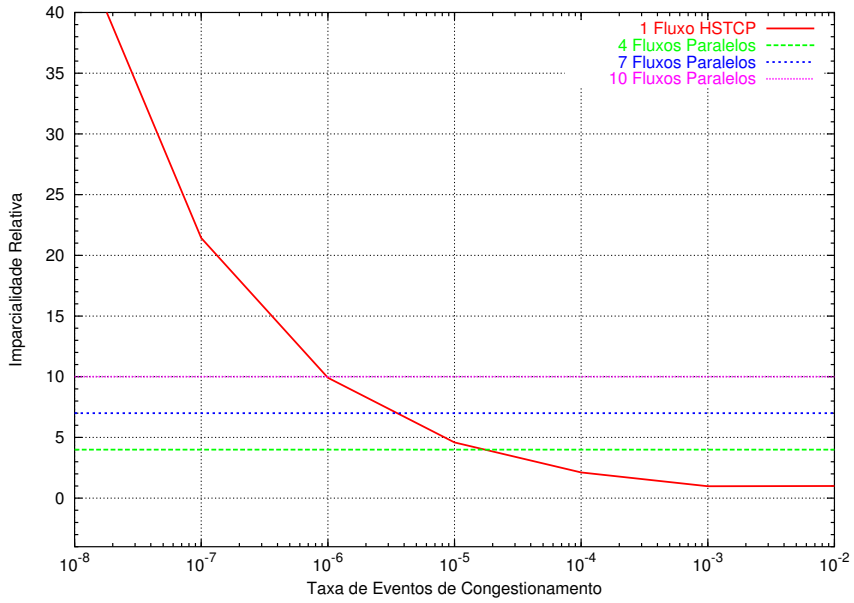


Figura 5.29: Imparcialidade Relativa por Fluxo Teórica - Transferência por Fluxos Paralelos

5.9 Fluxos Paralelos em Condição de Enlace com Perdas

O foco deste conjunto de experimentos foi observar o impacto do uso de fluxos paralelos sobre fluxos REGTCP de longa duração e compará-lo com o impacto do uso de HSTCP sobre o mesmo conjunto de fluxos REGTCP, quando ambos fossem submetidos a perdas sistêmicas. Nós utilizamos o modelo de perdas do simulador para simular perdas no enlace gargalo. Este modelo de perdas foi configurado para descartar pacotes com uma taxa média definida. A taxa de perdas usada foi 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3} e 10^{-2}. Nós utilizamos dois conjuntos de fluxos para desenvolver este experimento. O primeiro conjunto continha 10 fluxos REGTCP (representando os fluxos de longa duração) e também 1, 4, 7, 10, 20 ou 30 fluxos paralelos. O segundo conjunto foi formado pelos mesmos 10 fluxos REGTCP do primeiro conjunto e um fluxo HSTCP. Cada conjunto de fluxos rodou com as políticas de enfileiramento RED e DT nos roteadores. Cada simulação rodou por trezentos segundos e cada uma foi repetida dez vezes. A linha cruzando os pontos representa a mediana destas dez simulações. Cada repetição diferiu da outra apenas pelo número aleatório gerado no início de cada simulação. Os resultados apresentam o desempenho das métricas selecionadas para este experimento.

Os dois gráficos na Figura 5.30 apresentam a utilização de enlace agregada para os 10 fluxos REGTCP de longa duração, quando RED e DT são empregados, respectivamente. Também está presente nestes gráficos o desempenho de 10 fluxos REGTCP de longa duração quando não existem fluxos paralelos presentes.

Estes gráficos nos mostram que o desempenho de 10 fluxos de TCP Padrão de longa duração não é impactado pelo uso de fluxos paralelos até que a utilização completa do enlace seja alcançada. O mesmo ocorre quando um fluxo HSTCP é usado.

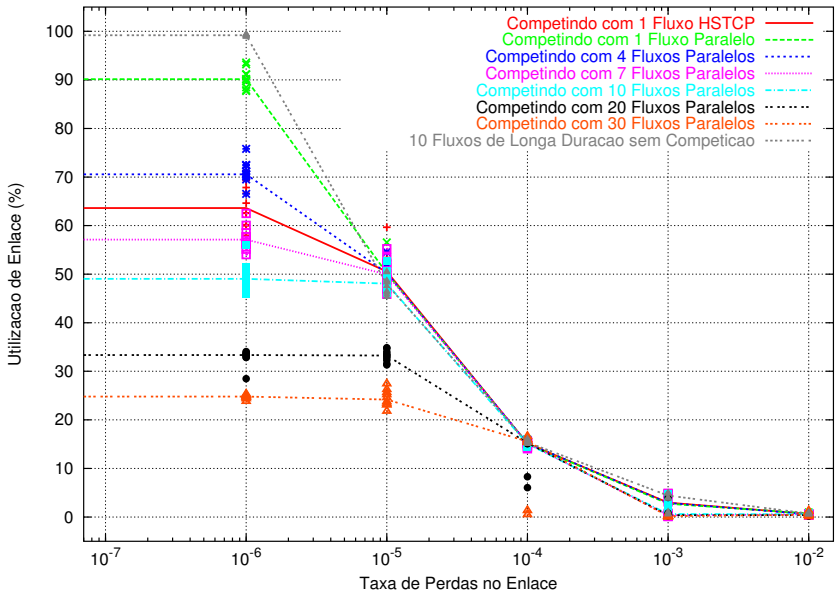
A informação descrevendo o impacto do uso de fluxos paralelos é complementada pelos gráficos da Figura 5.31. Eles mostram o desempenho dos fluxos paralelos e HSTCP dentro deste contexto.

A informação importante disponível é que, após o limite da banda ter sido atingido, a utilização do enlace permanece constante, de acordo com o número de fluxos paralelos presentes. O mesmo ocorre com o fluxo HSTCP. Outro importante aspecto apresentado aqui é que, para uma taxa de perdas superior a 10^{-3}, o desempenho é ruim para todos os esquemas.

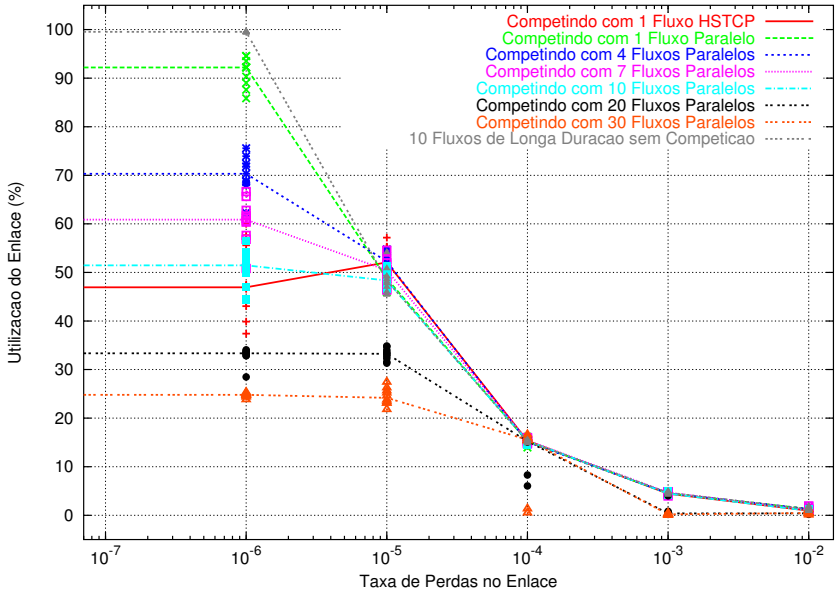
Os próximos dois gráficos da Figura 5.32 apresentam a taxa de eventos de congestionamento para o primeiro e segundo conjunto de de fluxos, quando o RED e DT são usados.

Finalmente nós apresentamos a imparcialidade relativa por fluxo. A intenção aqui é mostrar o nível de competição que uma transmissão por fluxos paralelos representa para um único fluxo de TCP Padrão de longa duração. A quantidade de banda utilizada pela transmissão por fluxos paralelos é dividida pela quantidade de banda utilizada por um dos 10 fluxos de longa duração. O mesmo procedimento é usado para o caso da transmissão usando um fluxo HSTCP. Os resultados estão apresentados na Figura 5.33.

Fica claro que, quando fluxos paralelos são empregados, a imparcialidade relativa é quase constante sobre uma ampla faixa de taxa de perdas no enlace. Este comportamento somente muda quando existe uma pesada taxa de perdas de pacotes. Em contraste, a imparcialidade relativa, quando HSTCP é usado, não é constante e tem uma ampla faixa de valores.

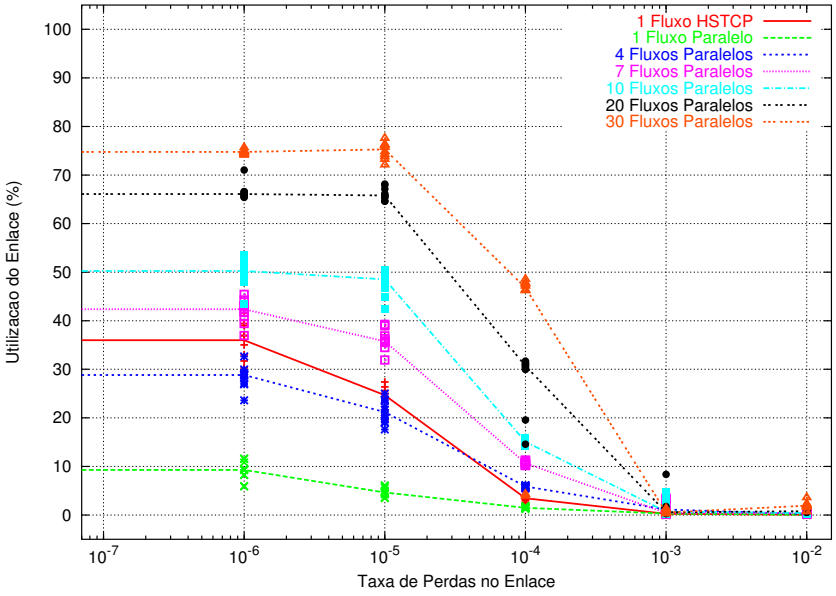


(a) RED

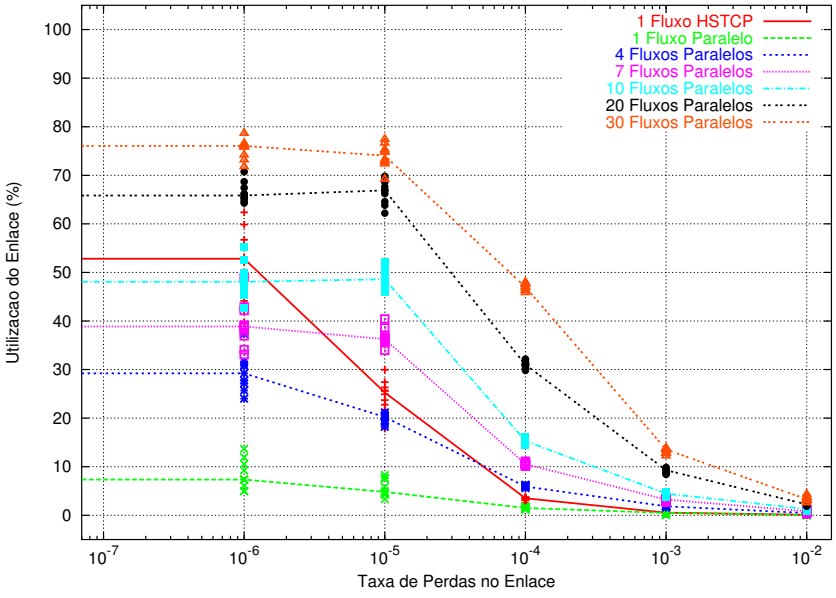


(b) DT

Figura 5.30: Utilização de Enlace Agregada para 10 fluxos REGTCP de Longa Duração - Fluxos Paralelos em Condição de Enlace com Perdas

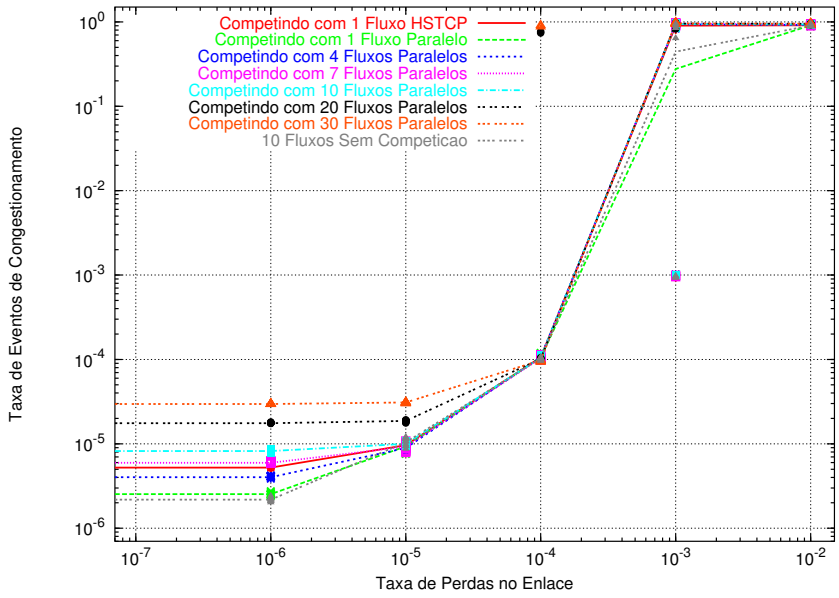


(a) RED

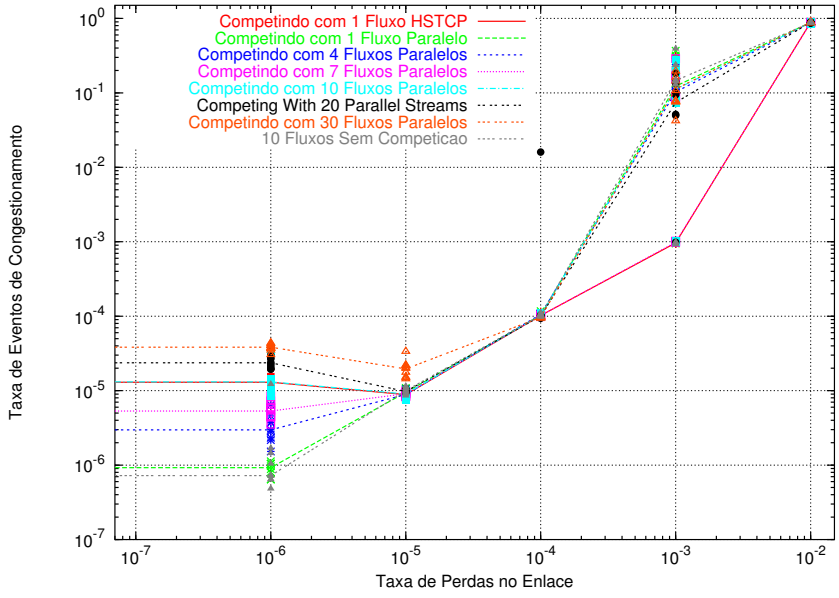


(b) DT

Figura 5.31: Utilização do Enlace Agregada dos Fluxos Paralelos Competidores - Fluxos Paralelos em Condição de Enlace com Perdas

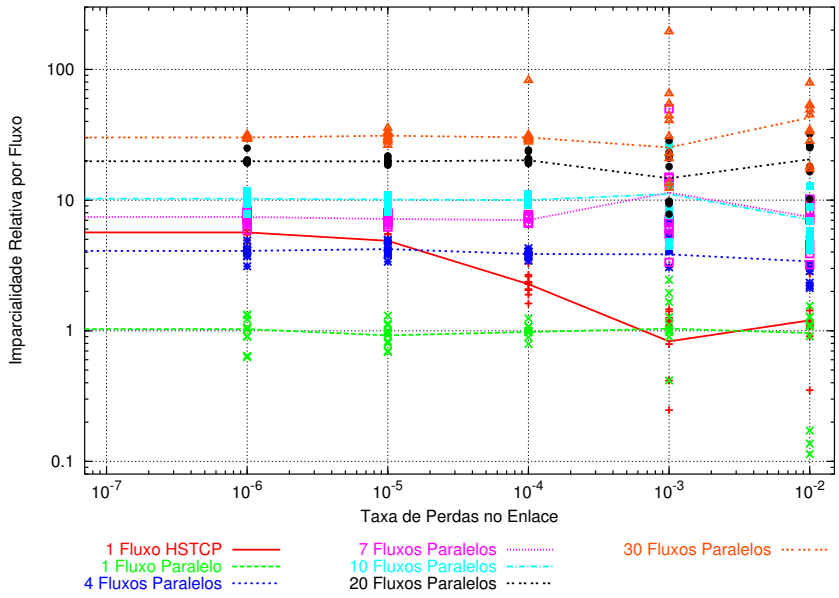


(a) RED

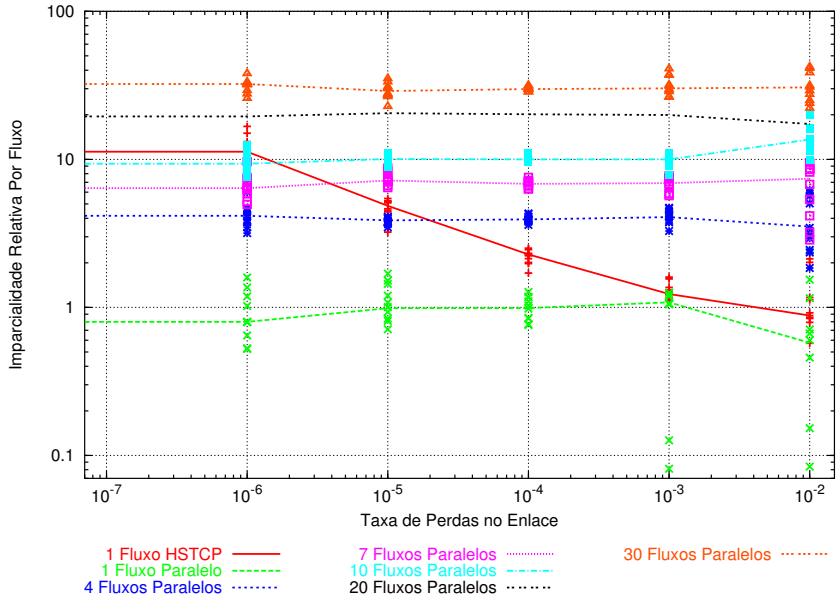


(b) DT

Figura 5.32: Taxa de Perdas no Enlace - Fluxos Paralelos em Condição de Enlace com Perdas



(a) RED



(b) DT

Figura 5.33: Imparcialidade Relativa por Fluxo - Fluxos Paralelos em Condição de Enlace com Perdas

5.10 Fluxos Paralelos em Condição de Tráfego em Rajadas

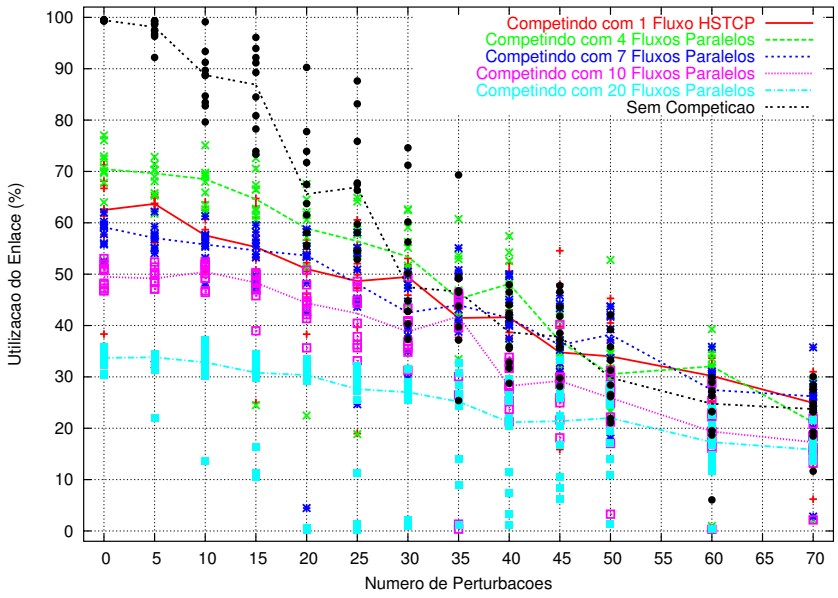
O objetivo deste conjunto de experimentos foi observar o impacto do uso de Fluxos Paralelos em fluxos REGTCP de longa duração e compará-lo com o impacto do HSTCP sobre o mesmo conjunto de fluxos, quando ambos são submetidos a tráfego em rajadas. O tráfego em rajada foi composto por fluxos de TCP Padrão de curta duração que duram apenas poucos segundos, de forma a só rodarem durante a fase de Partida Lenta. Como a fase de Partida Lenta possui um crescimento exponencial, os fluxos em rajada têm um considerável impacto nos fluxos de longa duração. Nós usamos 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60 e 70 fluxos em rajada distribuídos randomicamente, com distribuição uniforme, durante todo o período de simulação.

Nós utilizamos dois conjuntos de fluxos para desenvolver este experimento. O primeiro conjunto continha 10 fluxos REGTCP (representando os fluxos de longa duração) e também 1, 4, 7, 10 ou 20 fluxos paralelos. O segundo conjunto foi formado pelos mesmos 10 fluxos REGTCP do primeiro conjunto e um fluxo HSTCP. Cada conjunto de fluxos foi simulado com as políticas de enfileiramento RED e DT nos roteadores. Cada simulação durou trezentos segundos e cada uma foi repetida dez vezes. A linha cruzando os pontos representa a mediana destas dez simulações. Cada repetição diferiu da outra apenas pelo número aleatório gerado no início de cada simulação. Os resultados apresentam o desempenho das métricas selecionadas para este experimento.

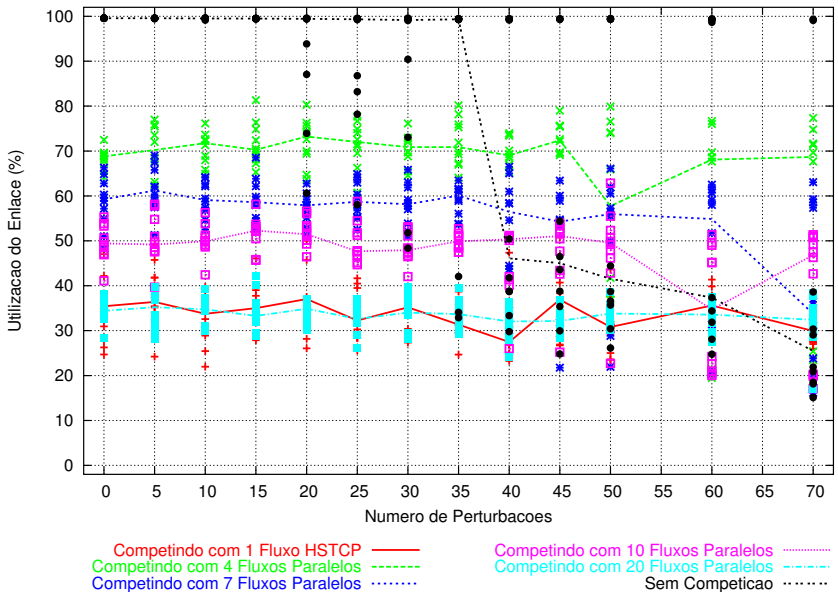
Os primeiros dois gráficos na Figura 5.34 apresentam a utilização do enlace para os 10 fluxos REGTCP de longa duração, quando RED e DT são empregados, respectivamente. Também está presente nestes gráficos o desempenho de 10 fluxos REGTCP de longa duração quando não existem fluxos paralelos competindo pela banda do enlace.

Os gráficos na Figura 5.35 completam as informações referentes ao impacto do uso de fluxos paralelos. Eles mostram o desempenho dos fluxos paralelos e do fluxo HSTCP dentro da condição de tráfego em rajadas.

Dois fatos são revelados através destes gráficos. O primeiro é que o desempenho dos fluxos paralelos tende a decrescer quando o número de perturbações aumenta. Isto é mais evidente quando RED é empregado do que quando DT é usado. O segundo fato é que o desempenho do HSTCP é muito menos sensível a este tipo de ambiente,

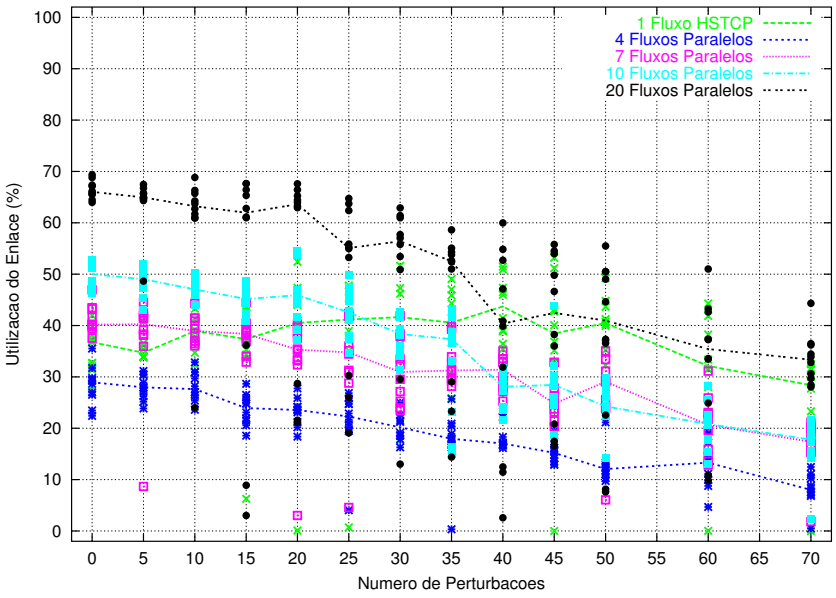


(a) RED

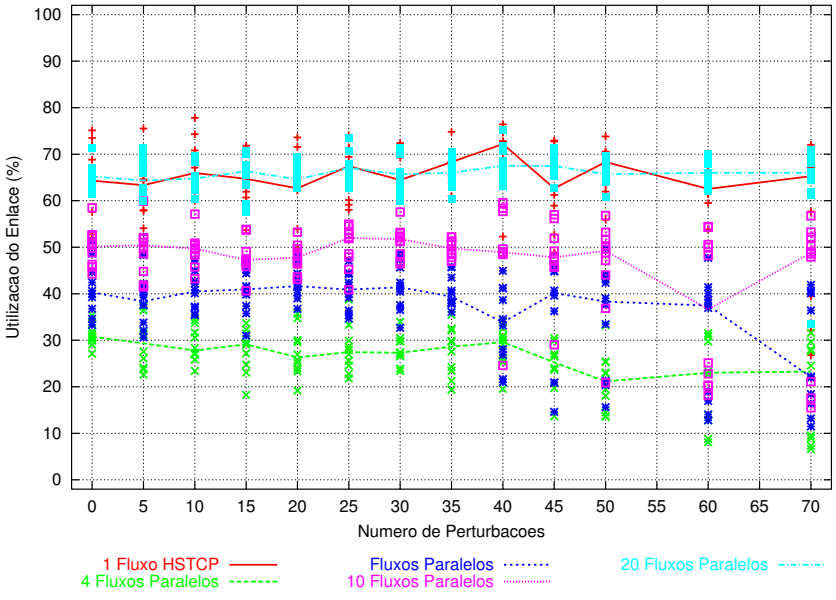


(b) DT

Figura 5.34: Utilização do Enlace Agregada para 10 fluxos REGTCP de Longa Duração - Fluxos Paralelos em Condição de Tráfego em Rajadas



(a) RED



(b) DT

Figura 5.35: Utilização do Enlace Agregada dos Fluxos Paralelos Competidores - Fluxos Paralelos em Condição de Tráfego em Rajadas

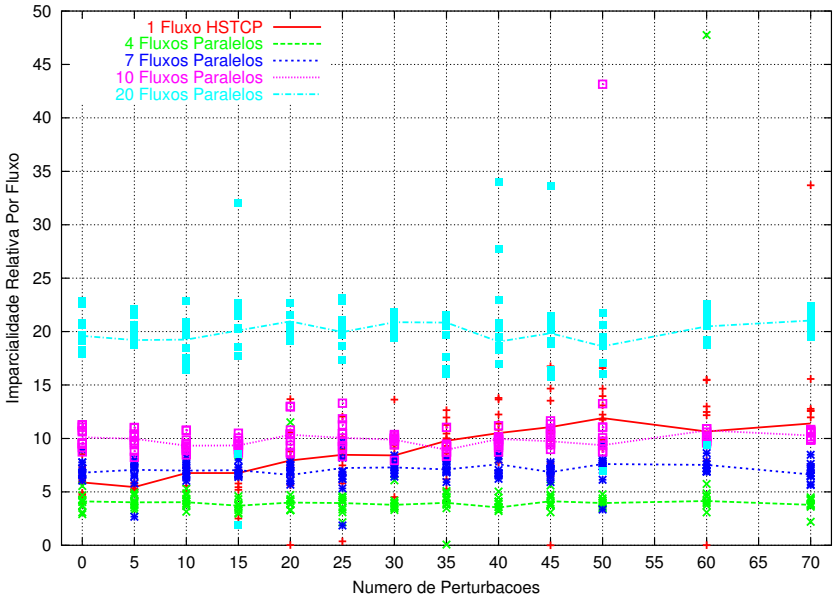
chegando mesmo a melhorar quando o número de perturbações aumenta.

A seguir nós apresentamos os resultados da imparcialidade relativa por fluxo. A intenção é mostrar o que representa uma transmissão por fluxos paralelos em termos de competição para um único fluxo de TCP Padrão de longa duração, quando ambos estão sujeitos ao tráfego em rajadas. A quantidade de banda no enlace usada pelos fluxos paralelos é dividida pela quantidade de banda no enlace usada por um dos 10 fluxos de longa duração. O mesmo procedimento é usado no caso da transmissão usando um fluxo HSTCP. Os resultados são apresentados na Figura 5.36.

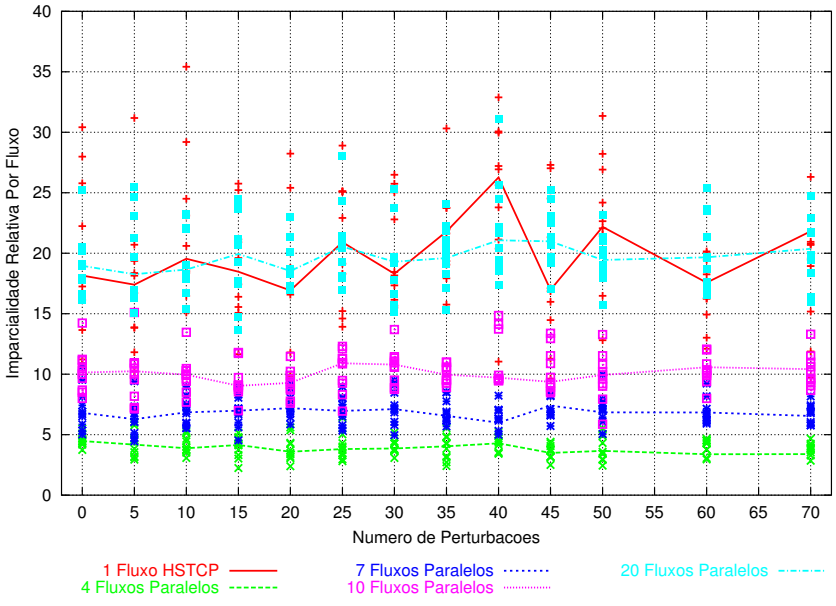
Nós observamos nestes gráficos que quando RED é usado a imparcialidade relativa com um fluxo HSTCP aumenta quando o número de perturbações aumenta, mas o mesmo comportamento não é claro quando DT é empregado. Em ambos os casos, a razão entre a banda usada pelo fluxo HSTCP e a banda usada por um dos 10 fluxos de longa duração pode espalhar-se sobre por uma grande faixa de valores.

A banda roubada dos fluxos TCP Padrão de longa duração quando eles são empregados conjuntamente com o fluxo HSTCP e os fluxos paralelos é apresentada na Figura 5.37.

A mensagem importante contida nestes gráficos é que a quantidade de banda roubada dos 10 fluxos TCP de longa duração diminui quando o número de perturbações aumenta, independente do tipo de gerenciamento de enfileiramento do roteador e independente do esquema de transferência volumosa de dados usada.

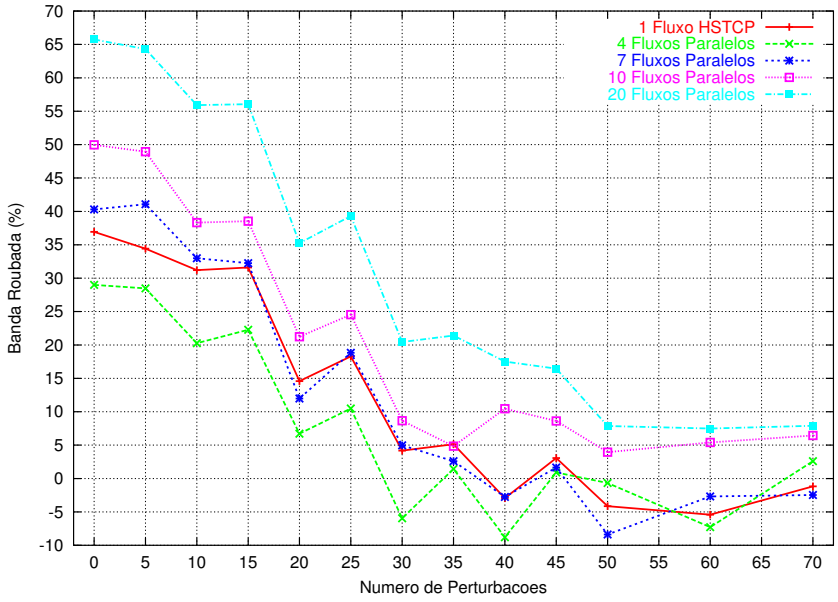


(a) RED

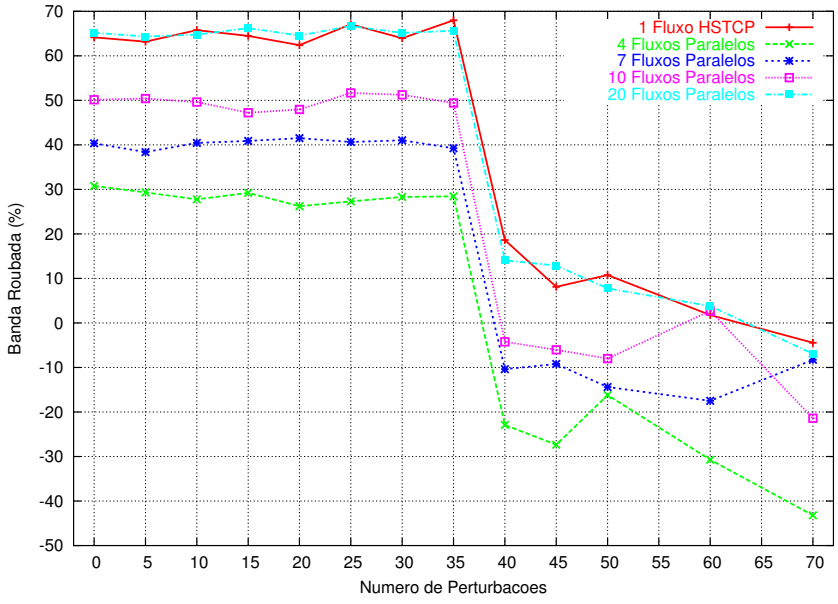


(b) DT

Figura 5.36: Imparcialidade Relativa por Fluxo - Fluxos Paralelos em Condição de Tráfego em Rajada



(a) RED



(b) DT

Figura 5.37: Banda Roubada - Imparcialidade Relativa por Fluxo - Fluxos Paralelos em Condição de Tráfego em Rajada

5.11 Partida Lenta

Muito embora este não seja um ponto central em nosso trabalho, nós desenvolvemos um experimento tendo como foco a fase de Partida Lenta, usando 1 fluxo HSTCP. A razão para isto foi procurar entender o efeito que alguns parâmetros da Partida Lenta têm sobre o estado estacionário de uma transmissão. Nós usamos dois algoritmos para alcançar este objetivo. O primeiro foi o algoritmo padrão de Partida Lenta e o segundo foi o algoritmo modificado de Partida Lenta (*Limited Slow-Start*), proposto em [17], para grandes janelas de congestionamento. O parâmetro MAX_SSTHRESH da Partida Lenta modificada foi configurado para 10, 100 e 1000. Cada algoritmo foi simulado uma única vez por 100 segundos de simulação, usando os parâmetros padrão para banda e atraso do enlace.

A Figura 5.38 apresenta a evolução da janela de congestionamento usando-se os algoritmos padrão e modificado da Partida Lenta. A Figura 5.39 apresenta a evolução do número de seqüência para a Partida Lenta padrão e modificada da mesma simulação. Finalmente a Tabela 5.1 apresenta a quantidade de pacotes descartados na simulação da Figura 5.38, na primeira metade da simulação.

Este experimento mostra o efeito das perdas de pacote na Partida Lenta. O crescimento do número de seqüência de um fluxo HSTCP é bem irregular até quase no sexto segundo para o algoritmo de Partida Lenta Padrão, como visto na Figura 5.39. Podemos observar também que o tamanho de janela no qual ocorreu a troca de fases foi muito elevado, conforme ilustrado na Tabela 5.1.

MAX-SSTHRESH	PCTS PERD/MARC	TEMPO DE TROCA	CWND NA MUDANÇA
0	27274	1.30	33032.00
10	1	63.75	9919.98
100	1	20.75	9979.66
1000	58	3.55	12726.20

Tabela 5.1: Comparação da Partida Lenta

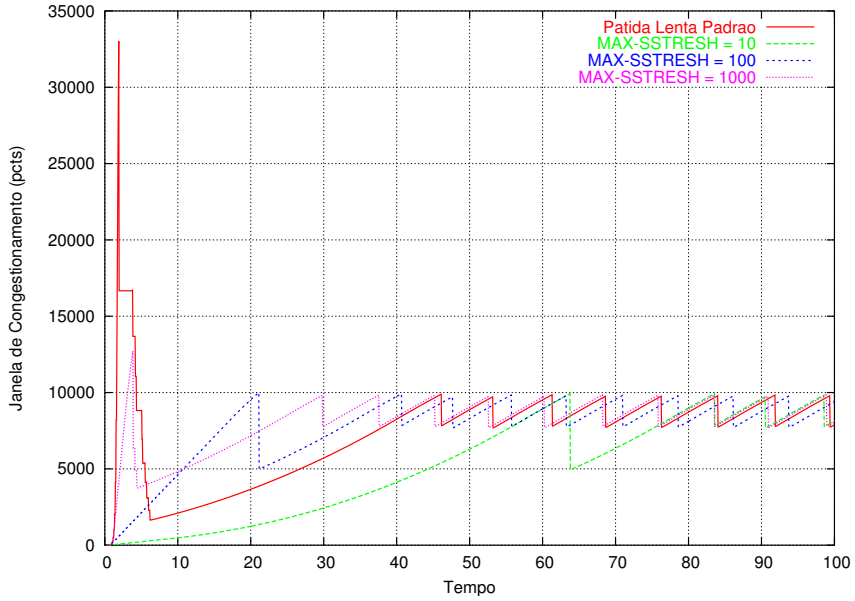


Figura 5.38: Variação da Partida Lenta com MAX-SSTHRESH

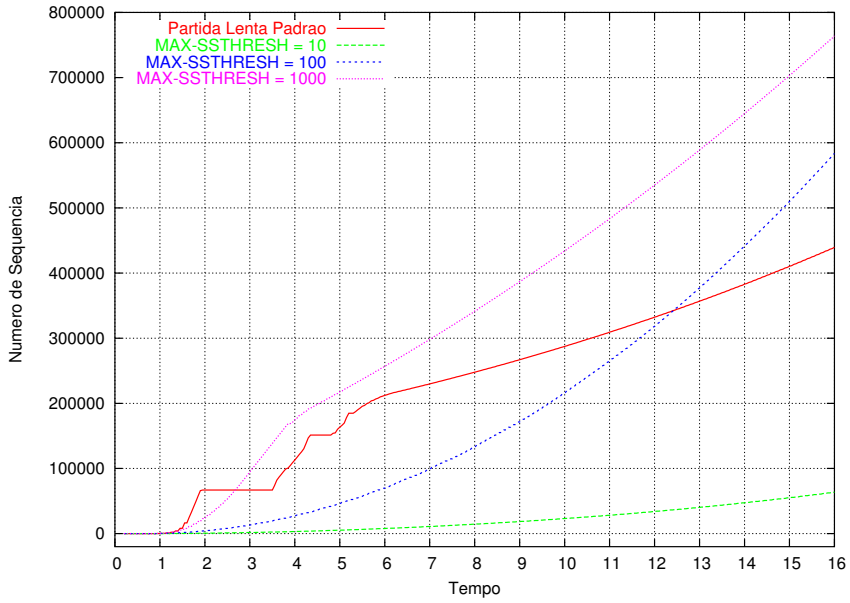


Figura 5.39: Efeito da Perda de Pacotes na Partida Lenta

Capítulo 6

Discussão dos Resultados

Este capítulo contém a discussão relativa aos resultados apresentados no capítulo anterior. As principais constatações dos resultados são analisadas e as questões propostas neste trabalho são discutidas aqui.

6.1 Organização dos Resultados

Os resultados estão apresentados da forma a seguir. Primeiro, nós apresentamos a comparação do desempenho do HSTCP e REGTCP para os ambientes de rede definidos nestes experimentos. É de especial interesse o desempenho do HSTCP nas situações onde REGTCP apresenta um fraco desempenho.

A segunda parte é dedicada ao entendimento das interações entre os fluxos do HSTCP e REGTCP. As questões relativas à imparcialidade são apresentadas e discutidas nesta parte, bem como uma análise da possibilidade do emprego do TCP de Alta Velocidade em conjunto com o TCP Padrão.

A terceira parte aborda a discussão sobre o impacto do tipo do gerenciamento de enfileiramento do roteador. O gerenciamento de *buffer* tem uma influência importante nas métricas de desempenho de ambos os protocolos, assim como em sua interação.

A quarta parte apresenta um exame do TCP de Alta Velocidade como método candidato para transferência volumosa de dados. Nós exploramos quantos fluxos REGTCP

um único fluxo HSTCP pode substituir e comparamos o desempenho do HSTCP com um outro método de transferência volumosa, em ambientes de rede diferentes.

A parte final aborda algumas outras questões envolvendo a Partida Lenta e a implementação do TCP, que surgiram durante o desenvolvimento deste estudo, mas que não estão diretamente relacionadas com o assunto deste estudo.

6.2 Questões sobre o Emprego do TCP de Alta Velocidade

6.2.1 Comportamento do TCP de Alta Velocidade em Situações de Baixo Desempenho do TCP Padrão

Condição Ideal

O ambiente usado para os testes na primeira condição de rede não impõe restrições a que os fluxos REGTCP atinjam a máxima utilização do enlace. O tempo para o experimento executar é longo o suficiente para esperar um bom desempenho. Muito embora isto seja verdade, um pequeno número de fluxos REGTCP é incapaz de usar completamente a banda disponível. Com menos de 10 fluxos, o conjunto de fluxos REGTCP não usa a banda totalmente neste ambiente de rede. Somente com um número maior de fluxos REGTCP é que a capacidade do enlace é alcançada. A Figura 5.2 apresenta esta situação. Se houver perdas sistêmicas no caminho da conexão, a situação piora, como pode ser visto na Figura 5.21.

A razão para este fraco desempenho está relacionada com o produto banda atraso do enlace gargalo e com o aumento conservador do CWND na fase de Prevenção de Congestionamento do REGTCP. Dado que, para cada segmento reconhecido durante a fase de Prevenção de Congestionamento, o CWND é aumentado em 1/CWND e o intervalo entre cada aumento é de 1 RTT, a evolução da janela de congestionamento é lenta. Este crescimento lento deixa o enlace com um baixo nível de utilização durante um período significativo de tempo. Esta situação é apresentada na Figura 5.1(a). Muito embora o REGTCP irá eventualmente alcançar o limite da banda, fica claro que, durante

uma grande porção de tempo, o enlace estará subutilizado.

É importante apontar que outras restrições a que o TCP alcance uma alta taxa de transferência, tais como *buffer* TCP limitado, reduzida capacidade de *buffer* de rede e grande perda de pacote na Partida Lenta, não estão presentes neste caso.

Em velocidades de Gbps, como a usada neste cenário de rede, a taxa de transferência do REGTCP é limitado pela latência, ao invés de ser limitado por banda. A latência é causada pela velocidade de propagação do sinal no meio físico e não pode ser diminuída. O outro possível limite é determinado pelo tamanho da janela do TCP [56]. Conforme visto aqui, o protocolo do TCP Padrão limita o desempenho através de sua taxa de aumento da janela de congestionamento lenta. Esta resposta dinâmica lenta incorrerá em outros problemas de desempenho que serão analisados mais tarde.

Ao contrário do REGTCP, poucos fluxos HSTCP são suficientes para encher o canal de dados, como pode ser visto na Figure 5.2. Esta é uma conseqüência direta do HSTCP ter uma taxa de incremento da CWND mais alta que a taxa de incremento que o REGTCP durante a fase de Prevenção de Congestionamento. Como previamente apresentado no Capítulo 3, o parâmetro usado para incrementar a CWND é função da janela de congestionamento atual. Quanto maior for a janela de congestionamento, maior será este fator.

Conforme visto no capítulo sobre os resultados, o HSTCP faz melhor uso de uma banda elevada, quando existe um atraso alto. Ele evita um longo período de crescimento exploratório da fase de Prevenção de Congestionamento e atinge um nível elevado de utilização em estado estacionário, num período curto de tempo.

A diferença de comportamento dos dois protocolos também é visível na Figura 5.3(a). Há uma clara diferença entre a taxa de eventos de congestionamento produzida por cada protocolo na mesma situação. Esta diferença pode ser entendida como diferença de agressividade. Como a CWND do HSTCP aumenta mais rapidamente que a CWND do REGTCP, a probabilidade de que um dos fluxos HSTCP alcance o limite da banda e conseqüentemente produza um evento de congestionamento, é maior do que para os fluxos REGTCP. Mesmo com um grande número de fluxos (30 ou 40) esta característica está presente. Este fato também sugere que, como o REGTCP é lento para ajustar sua janela de congestionamento em um ambiente de atraso elevado, quando um evento de congestionamento ocorre, a probabilidade de que um dos seus fluxos atinja o

limite da banda é menor que o mesmo ocorra com fluxos HSTCP.

É interessante observar que a taxa de eventos de congestionamento para um fluxo HSTCP não é inferior a 10^{-6}, como pode ser observado na Figura 5.3(a). Isto ocorre devido ao fato de que neste ponto um fluxo HSTCP atinge a taxa de transmissão máxima para as condições deste experimento. Este limite era previsto, bastando observar o ponto de cruzamento da linha da função resposta teórica do HSTCP e a linha de banda do enlace na Figura 3.1.

Condição de Enlace com Perdas

No segundo experimento, nós estudamos o impacto da perda de pacotes não causada pelo transbordamento do *buffer* do roteador, mas por uma transmissão falha, também chamada de perda sistêmica. O entendimento dominante é que perdas de pacote são causadas exclusivamente pelo descarte de pacotes nos roteadores, sendo isto interpretado pelo TCP como uma indicação de congestionamento na rede entre o transmissor e o receptor. Todavia, perdas de pacote podem ter outras fontes. Em [57], os autores mostram que entre 1 pacote em 1.100 e 1 pacote em 32.000 falham na conferência do TCP *checksum* e o pacote é descartado, mesmo quando o *CRC checksum* no nível de enlace passa.

Perdas de pacote podem também ser devidas a outros fatores randomicos além de congestionamento na rede, tais como falhas intermitentes de *hardware* [2]. Um exemplo de fonte de perda de pacotes não devida a congestionamento é descrita em [10], no qual uma grande quantidade de perdas de pacote em modens a cabo era causada por um defeito de *hardware*. Os autores em [46] encontraram que o descarte de células ATM, devido a problemas de *hardware*, limitava o desempenho do TCP em enlaces OC-12. Perdas não relacionadas a transbordamento de *buffer* também acontecem em conexões por satélite e outras formas de comunicações sem fio.

Dada esta evidência, a crença em que as perdas de pacotes são causadas somente por transbordamento de fila, não pode ser sustentada. Existem muitas partes de equipamento de rede e de transporte envolvidas em uma transmissão de longa distância, que um erro ou defeito em qualquer uma de suas partes pode corromper ou descartar um pacote.

Com este alto potencial para perdas sistêmicas, este experimento estudou o efeito de vários níveis de taxa de perdas de pacote nos dois tipos de TCP.

A utilização de enlace pelo conjunto contendo apenas fluxos REGTCP apresentou uma perda de desempenho impressionante quando a taxa de perdas no enlace aumentava. Isto impediu que o REGTCP fizesse um uso razoável da banda disponível, como ilustrado na Figura 5.8. Quando a taxa de perdas de pacote é maior que 10^{-6}, o REGTCP é incapaz de usar totalmente o enlace. Esta situação ocorre para o HSTCP apenas após 10^{-5}. O desempenho do conjunto de fluxos usando HSTCP ainda é melhor que o conjunto de fluxos usando REGTCP na faixa entre 10^{-6} e 10^{-3}. A taxa de perdas no enlace de 10^{-3} é o ponto configurado como Low_Window nos parâmetros HSTCP (uma janela de congestionamento de 31 pacotes representa uma taxa de eventos de congestionamento de aproximadamente 0.001). Acima desta taxa, o desempenho dos fluxos HSTCP é equivalente aos fluxos REGTCP, por conta do funcionamento igual dos protocolos.

É interessante notar que a taxa de eventos de congestionamento para os fluxos HSTCP não é menor que 10^{-5}, como pode ser visto na Figura 5.9(a). Em torno de 10^{-4} existe um ponto de inflexão na taxa de eventos de congestionamento. A razão para esta inflexão é que a influência das perdas no enlace fica menor do que a influência do congestionamento real induzido pelo limite na banda disponível. No caso onde o gerenciamento de enfileiramento do roteador RED é usado, o número de eventos de congestionamento produzidos por pacotes com ECN marcado foi maior que os eventos de congestionamento produzidos por perdas de pacote.

A Figura 5.8 também mostra que uma taxa de perdas no enlace entre 10^{-5} e 10^{-4} impede que os fluxos REGTCP façam um uso razoável da banda disponível no enlace (menos de 50% neste caso). Nesta mesma faixa, os fluxos HSTCP foram capazes de usar quase o dobro da banda usada pelos fluxos REGTCP, com os parâmetros usados na configuração dos fluxos HSTCP.

Uma outra maneira de entender o desempenho do HSTCP em relação ao REGTCP é observar a utilização do enlace quando uma taxa fixa de perdas no enlace é configurada, e é variado o número de fluxos. A Figura 5.21 mostra que o protocolo HSTCP precisa de apenas 6 fluxos para atingir plena utilização do enlace, enquanto que o protocolo REGTCP apenas alcança este desempenho com 20 ou mais fluxos. Esta situação é próxima da realidade, porque é difícil encontrar um caminho de rede sem perdas sistê-

micas.

A mesma inflexão presente na Figura 5.9(a) está ocorre também na Figura 5.22(a). O conjunto de fluxos REGTCP tem sua inflexão na taxa de eventos de congestionamento apenas após 30 fluxos, muito além dos 2 fluxos necessários no caso do HSTCP.

Condição de Tráfego em Rajadas

O tráfego de rede pode apresentar característica de rajada, devido a inerente característica de auto-similaridade [40]. Esta característica é de particular importância em nossa análise.

Em uma rede típica, o TCP otimiza a sua taxa de envio durante a Partida Lenta através da liberação de um número crescente de rajadas (ou janelas) de pacotes, uma rajada por tempo de percurso, para o receptor, até que alcance seu tamanho de janela máximo. Neste ponto ele alcança a capacidade plena da rede. Todavia em uma rede com um alto produto banda atraso, a janela de congestionamento máxima do TCP pode ser maior que a capacidade de fila de alguns dos roteadores intermediários. Grandes janelas sobrecarregam as filas dos roteadores e eles começam a descartar pacotes.

Este comportamento do TCP tem diversos efeitos nos elementos de rede, bem como no tráfego proveniente de outras fontes, presentes no mesmo enlace:

- Aumento no atraso de fila: A característica de rajada do tráfego leva ao enfileiramento dos pacotes nos roteadores intermediários. Entretanto, atrasos de *bufferização* podem também depender do nível de congestionamento, enfileiramento e políticas de sincronização. Grandes filas nos roteadores podem introduzir atrasos adicionais nos fluxos TCP e aumentar seu tempo de percurso.
- *Jitter* (variação no atraso): Em algumas redes, a variação no atraso decorre principalmente do enfileiramento de tráfego em rajadas.
- Tornar outro tráfego também com característica de rajada: A característica de rajada de um fluxo TCP pode induzir com que outros fluxos TCP fiquem em rajada quando eles compartilham a mesma fila em roteadores intermediários. O efeito pode ser catastrófico se todas as rajadas ficarem sincronizadas. O efeito da *sincronização global*, descrito em [63], pode fazer com que as perdas de pacotes devido

ao transbordamento de *buffer* fiquem sincronizadas, fazendo com todos os fluxos TCP retraiam-se simultaneamente e a banda fique subutilizada.

- Baixas taxas de transferência: A característica de rajada do TCP resulta em pacotes descartados surgindo a partir do transbordamento das filas ou num aumento de tempo de percurso devido a atrasos de enfileiramento. Isto conduz a uma baixa taxa de transferência para os fluxos TCP. Em conjunto com outros efeitos colaterais, como a sincronização na evolução da janela, pode gastar banda e também produzir perdas de pacote em rajadas.

- Compartilhamento injusto: A característica de rajada do TCP pode resultar numa competição injusta de tráfego entre os fluxos, causado por gargalo no enfileiramento.

O efeito de se ter tráfego em rajada competindo pelo enlace foi significativo em todos os conjuntos de fluxo.

O conjunto de fluxos REGTCP apresentou um declínio continuado em sua utilização do enlace, quando o número de perturbações em rajada aumentava. Isto pode ser visto na Figura 5.13(a). A explicação é que a janela de congestionamento dos fluxos REGTCP diminui pela metade cada vez que um evento de congestionamento ocorre e ela não recupera a taxa de transferência anterior devido ao lento incremento na janela de congestionamento. Quando existe uma alta banda no enlace, isto conduz a uma baixa utilização da banda disponível.

Esta deficiência na utilização de enlace não é devida a que as perturbações estejam usando uma grande porção de banda no enlace. Ao invés disto, ela é devida à característica de rajada das perturbações. É possível averiguar que os fluxos de perturbação usaram menos de 10% da capacidade do enlace, quando a utilização do enlace pelos fluxos REGTCP caiu em torno de 70%.

O número de eventos de congestionamento para o conjunto de fluxos REGTCP é baixo muito embora o número de perturbações possa ser alto. Isto acontece porque, como os fluxos REGTCP diminuem sua taxa de transferência, a probabilidade de que algum dos fluxos tenha uma pacote na fila é baixo e a maioria dos pacotes perdidos pertencerá às perturbações.

O conjunto de fluxos HSTCP também diminui a sua utilização do enlace na presença de perturbações em rajada, porém ele o faz suave e lentamente quando o número de perturbações aumenta, com pode ser visto na Figura 5.13(a). Fica claro que os fluxos HSTCP perdem menos banda que os fluxos REGTCP, portanto ele é mais resistente a este tipo de tráfego e recupera-se mais rapidamente desta interferência.

A quantidade de banda utilizada pelas perturbações quando competindo contra os fluxos HSTCP é levemente inferior à banda usada pelas perturbações quando competindo contra o conjunto de fluxos REGTCP, conforme visto na Figura 5.13(a). Isto sugere duas coisas. Primeiro, a quantidade de banda perdida pelos fluxos REGTCP novamente não é devido às perturbações estarem usando esta banda. Segundo, as perturbações em rajada têm uma competição mais dura contra os fluxos HSTCP para usar a banda disponível, ou visto de outra forma, os fluxos HSTCP resistem melhor a este tipo de perturbação.

6.2.2 Manutenção da Imparcialidade na Utilização do TCP de Alta Velocidade junto com o TCP Padrão

Os fluxos HSTCP apresentam um melhor desempenho que os fluxos REGTCP em enlaces de alta velocidade e longo atraso, conforme visto na seção anterior. É evidente que o HSTCP é mais agressivo que o REGTCP, o que contribui para este desempenho. Neste tópico, uma questão importante é levantada: quão prejudiciais são os fluxos HSTCP para o desempenho dos fluxos REGTCP quando ambos são empregados conjuntamente?

A imparcialidade é um ponto chave na aceitação de um novo protocolo ou solução numa rede de melhor esforço. Este aspecto suscitou preocupação na comunidade de redes no passado [15] e torna difícil o emprego e a coexistência com outros protocolos [25]. Se uma implementação de controle de congestionamento é muito mais agressiva no uso da banda que outras implementações, isto pode induzir os novos protocolos a serem mais agressivos também.

Um fluxo é *TCP-Amigável* se ele é responsivo a notificação de congestionamento e em estado estacionário não usa mais banda que um fluxo TCP conformante rodando em condições comparáveis de taxa de perda de pacotes, RTT e MTU. Como mencionado

anteriormente, de uma certa forma, o HSTCP não é *TCP-Amigável*, mas seu grau de compatibilidade muda de acordo com a taxa de perda de pacotes percebida pelo fluxo HSTCP.

Nós estudamos a imparcialidade em estado estacionário, que é importante para uma transferência volumosa de dados. Outras situações, onde a imparcialidade também é importante, tais como na Partida Lenta e em algumas condições transitórias, não serão analisadas aqui, sendo deixadas para trabalhos futuros.

Condição Ideal

Se não há interferência externa, a porção de banda usada pelos fluxos HSTCP é maior que a usada pelos fluxos REGTCP, quando ambos os tipos de fluxo competem pelo mesmo enlace. Esta diferença na utilização do enlace pode ser claramente vista na Figura 5.4(a). O enlace apresenta um elevado nível de utilização quando os dois tipos de fluxo são empregados conjuntamente. É notável que a quantidade de banda usada pelos fluxos HSTCP decresce quando o número total de fluxos aumenta. O oposto ocorre com os fluxos REGTCP. A razão para este comportamento é encontrada no fato que a atualização do crescimento da janela de congestionamento do HSTCP está ligada com o nível da taxa de eventos de congestionamento percebida pelos fluxos. Quanto maior for o número de fluxos competindo pela banda do enlace, mais eventos de congestionamento ocorrerão e menos agressivos serão os fluxos HSTCP. Com a diminuição da agressividade do HSTCP, os fluxos REGTCP terão mais oportunidade de usar a banda disponível.

O mesmo comportamento é observado na imparcialidade relativa, como visto na Figura 5.6. A desproporção entre as utilizações do enlace decresce quando o número de fluxos aumenta. Uma ampla faixa de valores existe quando existe um pequeno número de fluxos competindo pelo enlace. A razão para isto pode estar associada com a ocorrência de um evento de congestionamento bem na fase inicial que reduz significativamente o tamanho da janela de congestionamento do fluxo REGTCP. Como existem poucos fluxos REGTCP, isto contribui para uma utilização do enlace em geral baixa. O momento em que um evento de congestionamento ocorre irá definir a utilização do enlace pelo fluxo REGTCP e conseqüentemente a imparcialidade relativa. A diferença pode ser grande, mas não o suficiente para impedir que os fluxos REGTCP usem parte da banda

do enlace.

A Figura 5.5 apresenta a evolução da taxa de eventos de congestionamento e mostra o aumento na taxa de eventos de congestionamento quando o número de fluxos aumenta, estabelecendo uma relação entre eles.

A porcentagem do total da capacidade de banda que os fluxos HSTCP roubam dos fluxos REGTCP é apresentada na Figura 5.7. Novamente, o resultado da variação na taxa de eventos de congestionamento é visível nos resultados.

Os resultados apresentados anteriormente ressaltam duas distintas características do protocolo HSTCP. Ele é mais agressivo no uso da banda disponível, mas diminui sua agressividade quando a taxa de eventos de congestionamento aumenta. Esta adaptabilidade é muito interessante no contexto de enlaces de alta velocidade. Ela evita que o enlace fique ocioso devido à dinâmica lenta do TCP Padrão, mas não evita que mais fluxos do TCP Padrão obtenham uma porção razoável do enlace. Esta adaptabilidade é expressa na Figura 5.19(a). Neste experimento, também é possível ver como a banda é repartida quando existe apenas um fluxo HSTCP e o número de fluxos REGTCP competindo pelo enlace aumenta. A utilização do enlace pelo fluxo HSTCP retrai-se quando o número de fluxos REGTCP aumenta. A utilização do enlace total é mantida perto de 100%, o que significa que o recurso de banda é totalmente utilizado.

No mesmo gráfico, o ponto de intersecção das linhas de utilização do enlace pelo fluxo HSTCP e a linha de utilização do enlace pelos fluxos REGTCP representa o momento onde 1 fluxo HSTCP é equivalente a $(N-1)$ fluxos REGTCP, ou ambos utilizam o mesmo nível de banda. Para as condições de rede deste experimento N = 7, o que significa que 1 fluxo HSTCP é equivalente a 6 fluxos REGTCP.

Até aqui, nós apresentamos a interação entre fluxos HSTCP e fluxos REGTCP por um período de tempo de 300 segundos. É também importante ver sua interação ao longo de um período maior de tempo. A Figura 5.26 apresenta a evolução da janela de congestionamento de 1 fluxo HSTCP e de 1 fluxo REGTCP competindo entre si, por um período de tempo de uma hora. Fica claro que existe pouca diferença nesta interação. As janelas de congestionamento são mantidas em torno do mesmo nível durante todo o período. A ampla área ocupada pela linha do HSTCP representa a oscilação da janela de congestionamento do HSTCP.

Condição de Enlace com Perdas

Na condição de enlace com perdas, a diferença na utilização do enlace dos fluxos HSTCP e dos fluxos REGTCP decresce com o aumento do número de perdas, conforme é esperado, vide Figura 5.10(a), e a imparcialidade relativa diminui com o aumento da taxa de perdas no enlace, como é visto na Figura 5.11.

Uma importante questão a ser considerada é quanto da banda do enlace os fluxos HSTCP roubam dos fluxos REGTCP e onde isto acontece. A resposta está na Figura 5.12. A conclusão é que, com uma taxa de perdas no enlace maior que 10^{-4}, os fluxos REGTCP são limitados por perdas sistêmicas, ao invés de terem uma perda de desempenho devido ao uso do HSTCP. Com uma taxa de perdas inferior a 10^{-4}, os fluxos HSTCP começam a roubar banda dos fluxos REGTCP. Esta mudança ocorre devido a utilização do enlace estar próxima do limite físico do enlace, como pode ser observado na figura 5.10(a). Além deste ponto, os fluxos HSTCP estarão competindo diretamente com os fluxos REGTCP por mais banda. Este ponto de mudança pode ocorrer de acordo com a capacidade do enlace e com o número de fluxos competindo pelo enlace. Este resultado é semelhante ao encontrado na literatura [25].

Condição de Tráfego em Rajadas

Como mencionado anteriormente, o tráfego em rajadas prejudica o desempenho de ambos os conjuntos de fluxo. Entretanto, a maior diferença aqui está no desempenho dos fluxos REGTCP, como ilustrado na Figura 5.16(a). A queda na utilização do enlace agora é menor do que quando os fluxos REGTCP estão sofrendo as perturbações sozinhos, como pode ser visto na Figura 5.13(a). Uma possível explicação para este comportamento é que os fluxos REGTCP já possuem uma taxa de transferência baixa, porque estão competindo contra os fluxos HSTCP. A imparcialidade relativa para este experimento em particular é mantida relativamente constante, conforme visto na Figura 5.17.

Também é possível saber o quanto de banda é roubado pelos fluxos HSTCP dos fluxos REGTCP. A resposta é encontrada na Figura 5.18. Quando competindo contra os fluxos HSTCP em um ambiente de tráfego em rajadas, o desempenho dos fluxos REGTCP é relativamente constante. Isto produz o seguinte resultado. Quando o nível

de perturbação no enlace aumenta, a diferença entre os fluxos REGTCP competindo com outros fluxos REGTCP, e quando competindo contra os fluxos HSTCP, diminui. A conclusão é que o tráfego em rajadas tem pouca influência na quantidade de banda que os fluxos HSTCP roubam dos fluxos REGTCP e, portanto, também tem diminuta influência na imparcialidade.

6.2.3 O Efeito da Política de Enfileiramento do Roteador

O desempenho e a imparcialidade do HSTCP não pode ser completamente entendidos sem se identificar a influência que o esquema de gerenciamento de enfileiramento do roteador tem sobre eles.

Impacto do RED e DT no Desempenho do HSTCP

O esquema de gerenciamento de fila não afeta significativamente a utilização do enlace dos fluxos HSTCP na Condição Ideal. Eles apresentam resultados semelhantes, como indicado anteriormente.

Existe uma considerável diferença no nível da taxa de eventos de congestionamento do RED para o DT, conforme ilustrado na Figura 5.3(b). O uso do RED causa a redução no número de eventos de congestionamento necessários para controlar a taxa de envio do TCP. O ECN desempenha um importante papel notificando os transmissores TCP do congestionamento em formação de uma maneira mais efetiva do que através do descarte de pacotes [49].

Na condição de enlace com perdas, definida para este experimento, não existe diferença entre o uso do RED e do DT, como mencionado anteriormente. A razão simples para isto é que a fila do roteador não é realmente usada, porque o conjunto de fluxos HSTCP é limitado por perdas sistêmicas ao invés de perdas por congestionamento. Portanto, a quantidade de tráfego gerado não alcança a capacidade do enlace. O gerenciamento da fila somente está ativo para taxas de perda no enlace configuradas para valores inferiores a 10^{-5}, conforme a Figura 5.8.

O impacto do gerenciamento de enfileiramento do roteador é claro quando o

HSTCP é submetido a tráfego em rajadas. A utilização do enlace pelo conjunto de fluxos HSTCP decresce levemente com o RED, mas não é afetado pelas perturbações quando a política de filas no roteador DT é usada. Isto pode ser comprovado na Figura 5.13(b). A taxa constante de eventos de congestionamento do HSTCP, apresentada na Figura 5.14(b), também ilustra este fato. O RED diminui a propensão contra o tráfego em rajadas, aumentado os eventos de congestionamento do tráfego não em rajadas, como ilustrado na Figura 5.14(a). Este resultado também é encontrado na literatura [3].

Observando a Figura 5.14(b), também é possível perceber que os fluxos HSTCP parecem resistir melhor à sincronização global do que os fluxos REGTCP. A divisão dos resultados da taxa de eventos de congestionamento dos fluxos REGTCP em dois agrupamentos, parece indicar que, em um certo nível de perturbação, a sincronização global pode ser acionada. Uma vez que ela inicia, os fluxos de longa duração vão para a fase de Partida Lenta conjuntamente, tornando mais provável que a sincronização global seja sustentada pelas perturbações seguintes e que os fluxos REGTCP não consigam se recuperar.

Impacto do RED e DT na Imparcialidade

O padrão geral da imparcialidade relativa encontrada quando o RED é usado, também é seguido quando o gerenciamento de enfileiramento do roteador DT é empregado. A diferença está na maior quantidade de banda que os fluxos HSTCP tomam dos fluxos REGTCP. Na Condição Ideal, a razão atinge valores maiores que 20 vezes para o DT e possui uma alta variabilidade, como apresentado na Figura 5.6.

O uso do DT como gerenciamento de enfileiramento do roteador também muda a quantidade de banda roubada pelos fluxos HSTCP dos fluxos REGTCP, conforme visto na Figura 5.7. Muito embora a quantidade de banda roubada decresça quanto o número de fluxos aumenta, a distância entre as quantidades roubadas aumenta. Isto sugere que o RED está fazendo um melhor papel para evitar uma divisão injusta entre os fluxos na fila do roteador.

Não existem observações relevantes na Condição de Enlace com Perdas porque a fila não é usada na maior parte deste experimento.

Na Condição de Tráfego em Rajadas, a imparcialidade relativa é quase constante quando o DT é usado, mas parece aumentar levemente quando o RED é empregado. Este fato será mais explorado quando a capacidade do HSTCP de fazer uma transferência volumosa de dados for analisada. Usando-se o DT, os fluxos HSTCP adquirem entre 10 e 15 vezes mais porção de banda do que os fluxos REGTCP, como ilustrado na Figura 5.17.

A quantidade de banda roubada pelos fluxos HSTCP dos fluxos REGTCP também decresce com o incremento no número de perturbações quando DT é empregado, principalmente devido ao efeito que as pertubações em rajadas tem nos fluxos REGTCP. Entretanto, a quantidade de banda roubada ainda é superior do que a quantidade roubada com o gerenciamento de enfileiramento do roteador RED, conforme visto na Figura 5.18.

O experimento com Enlace com Taxa Constante de 10^{-5} apresenta uma nova situação para a análise da imparcialidade. A Figura 5.23(b) mostra que a porcentagem de banda usada por ambos os tipos de TCP é relativamente invariante com o número de fluxos usados. Isto é diferente do resultado encontrado quando não havia perdas no enlace, conforme visto na Figura 5.4(b). A pequena variação que acontece parece indicar que a diferença de utilização do enlace entre os fluxos HSTCP e fluxos REGTCP torna-se maior. Isto está claramente ilustrado na Figura 5.24 e na Figura 5.25. Uma investigação futura é necessária para explicar este comportamento.

O gerenciamento de enfileiramento do roteador também faz diferença no número fluxos REGTCP equivalentes a um fluxo HSTCP, quando ambos estão competindo pelo mesmo enlace. Na Figura 5.19(b), o ponto de equivalência para o gerenciamento de enfileiramento do roteador DT (o ponto onde 1 fluxo HSTCP consome a mesma quantidade de banda que N-1 fluxos REGTCP) é duas vezes maior que o valor encontrado quando RED é usado. O valor encontrado para o DT para as condições deste experimento foi de 13 fluxos REGTCP para 1 fluxo HSTCP. Este número ressalta novamente o impacto substancial que tipos diferentes de gerenciamento de enfileiramento podem ter no comportamento dos fluxos TCP e seu conseqüente impacto na imparcialidade observada quando os fluxos HSTCP e REGTCP são empregados conjuntamente.

6.2.4 O TCP de Alta Velocidade como Substituto para Outros Tipos de Transferência Volumosa de Dados

O compartilhamento de recursos é a força motriz por trás do desenvolvimento das redes de computadores. Ele permite que recursos escassos e dispendiosos sejam usados por pessoas dispersas geograficamente. Em certos tipos de comunidade, o recurso a ser compartilhado pode ser grandes quantidades de dados produzidos por experimentos, coleções de dados, ferramentas de visualização, e assim por diante. Este cenário é particularmente prevalente em comunidades científicas, tais como física de alta energia, clima, astronomia e ciências biológicas [34]. Grandes conjuntos de dados produzidos em uma localidade muitas vezes precisam ser analisados em colaboração com instituições ao redor do mundo.

É um enorme problema assegurar que os dados sejam distribuídos em tempo aceitável para seu processamento na Internet. Este problema tem forçado o projeto de novas técnicas para superar este desafio. Uma das técnicas atualmente em uso para realizar a transferência volumosa de dados é a utilização de fluxos TCP paralelos, como mencionado no Capítulo 2. Esta técnica é implementada através da divisão dos dados a serem transmitidos em N porções e da transferência em separado de cada porção em uma conexão TCP. Quando N conexões TCP estão rodando, será menos provável que cada fluxo paralelo venha a ser selecionado para ter seus pacotes descartados, e portanto, a quantidade agregada de banda potencial que precisará ir para uma fase de Prevenção de Congestionamento ou Partida Lenta prematura é reduzida.

Já foi demonstrado em [31] que a utilização do enlace é proporcional à probabilidade de perdas e ao produto banda atraso. O efeito de N fluxos paralelos é reduzir o produto banda atraso experimentado por um único fluxo por um fator N, porque todos os N fluxos compartilham a mesma banda.

Hacker e Athey [25] abordaram como o uso de conexões TCP paralelas aumenta a taxa de transferência agregada e como determinar o número de conexões TCP que são necessárias para maximizar a taxa de transferência, evitando o congestionamento na rede. Após o desenvolvimento de um modelo teórico e experimentos, eles concluíram que o uso de conexões TCP paralelas é equivalente a usar um grande MSS em uma única conexão, como o benefício de reduzir os efeitos negativos de perdas de pacotes randômicas. Eles também mencionaram que o número de conexões TCP paralelas não

pode ser arbitrariamente selecionado, porque, se o valor for muito alto, o fluxo agregado pode causar congestionamento e a taxa de transferência não será maximizada.

Em trabalho posterior, Hacker, Noble e Athey [26] propuseram um controle de congestionamento fracionário. Neste modelo, é informado individualmente a cada fluxo TCP que aumente a sua janela de congestionamento por somente um pacote a cada N pacotes de reconhecimento, mas diminua a sua janela da maneira normal. O controle de congestionamento fracionário pode ser usado para reduzir a agressividade dos fluxos paralelos em presença de congestionamento, mas preservar muito de sua efetividade na sua ausência. Eles propuseram este controle porque quando os fluxos paralelos competem com um único fluxo TCP, o primeiro rouba banda do último.

As principais vantagens e desvantagens do uso de fluxos paralelos estão resumidas e listadas abaixo:

VANTAGENS:

- a Partida Lenta é mais rápida porque o fluxo agregado cresce N vezes mais rápido;
- fluxos paralelos podem superar a limitação no tamanho máximo de *buffer*, como discutido no capítulo 2;
- sua recuperação é mais rápida quando comparada com um único fluxo TCP com uma grande janela de congestionamento, porque a recuperação de N janelas individuais é mais rápida que a recuperação de uma única e, se somente um fluxo dos N experimentar perdas, o decréscimo não será tão grande para o fluxo agregado

DESVANTAGENS:

- fluxos paralelos requerem suporte especial nas aplicações e conseqüentemente programas já existentes devem ser modificados;
- fluxos paralelos podem perder desempenho se a perda experimentada pelo fluxo agregado for devida a congestionamento;
- a seleção do número de fluxos paralelos é problemática, porque as condições de rede podem mudar durante o transcurso de uma transmissão e, uma condição previamente boa, pode tornar-se ruim posteriormente;

- fluxos paralelos podem ser injustos com outros fluxos TCP no compartilhamento dos mesmos recursos de rede

O emprego do TCP de Alta Velocidade apresenta vantagens e desvantagens comparado com o uso da técnica de fluxos paralelos:

VANTAGENS:

- o TCP de Alta Velocidade não requer mudanças nos programas aplicativos para usá-lo, mas apenas uma mudança na pilha TCP;
- sua adaptabilidade a taxas de perda variáveis acomoda melhor mudanças nas condições de rede, mesmo num enlace congestionado (muito embora o uso do controle de congestionamento fracionário melhore este ponto, ele ainda será dependente do número inicial de fluxos paralelos configurado para uma transmissão);
- não é necessário saber antecipadamente o número de fluxos a transmitir.

DESVANTAGENS:

- o TCP de Alta Velocidade tem a mesma limitação de tamanho máximo de *buffer* TCP para um único fluxo TCP;
- o TCP de Alta Velocidade tem um único laço de controle, ao invés de N, como no caso de fluxos paralelos. Isto implica que se em uma transmissão um fluxo HSTCP tiver problemas, toda a transmissão ficará prejudicada. Com fluxos paralelos isto não ocorre porque, no caso de um dos fluxos apresentar problemas, os demais ainda estarão realizando a transmissão.

Ambas as soluções têm em comum uma potencial parcialidade em relação a transmissões TCP concorrentes, quando elas compartilham um mesmo enlace congestionado e ambas apresentam uma clara melhoria em suas taxas de transmissão.

Apresentamos a seguir o desempenho teórico do HSTCP comparado com o desempenho teórico dos fluxos paralelos, bem como simulações comparando os aspectos de taxa de transferência e imparcialidade em condições de rede diferentes.

Desempenho e Imparcialidade do TCP de Alta Velocidade e Fluxos TCP Paralelos

Se o HSTCP é usado em uma rede com banda ainda disponível e existem perdas de pacote sistêmicas, as perdas de pacote experimentadas por um fluxo TCP irão determinar efetivamente a taxa de transferência máxima. A única alteração possível para mudar este limite é usar valores diferentes para os parâmetros do HSTCP (Low_Window, High_Window e High_P). Estes parâmetros definem a inclinação da função resposta. Após alcançar a capacidade do enlace, as perdas de pacote por congestionamento limitam um aumento maior da taxa de transferência. Os fluxos paralelos têm a sua taxa de transferência (e conseqüentemente a sua taxa de transferência agregada) determinada pela taxa de perdas de pacote experimentada em cada fluxo TCP. Quando o número de fluxos paralelos aumenta, as perdas de pacote percebidas por cada fluxo individual devem ser similares enquanto houver poucos pacotes sendo enfileirados nos roteadores. O desempenho de cada esquema é apresentado na Figura 5.28. Este gráfico apresenta somente o desempenho quando existem perdas de pacote sistêmicas e não inclui perdas devido a congestionamento.

Observando suas funções resposta, fica claro que um fluxo HSTCP pode fornecer uma taxa de transferência superior quando comparado a um único fluxo REGTCP ou mesmo comparado a fluxos paralelos, quando uma taxa de perdas de pacote sistêmica muito baixa está presente. Sob influência de uma taxa de perdas sistêmicas baixa, o HSTCP é um forte candidato para uma transferência volumosa de dados. Para ambientes com perdas de pacotes sistêmicas alta, a taxa de transferência do HSTCP fica perto da taxa de transferência de um fluxo REGTCP e nenhuma vantagem em particular existe em termos de taxa de transferência.

Se a capacidade do enlace é alcançada e perdas por congestionamento dominarem, não existe diferença entre os dois esquemas (veja a Figura 5.27), exceto a imparcialidade em relação a outro fluxo TCP. Este aspecto é explorado na discussão seguinte.

A imparcialidade relativa alcançada pelo uso de fluxos paralelos está diretamente relacionada com o número de fluxos usado (N) e independente da taxa de perdas de pacote experimentada pelos fluxos [18, 25]. Por outro lado, a imparcialidade relativa do HSTCP é uma função da janela de congestionamento, e conseqüentemente, uma função da taxa de eventos de congestionamento, como pode ser visto na Figura 5.29. Através deste gráfico, é possível ver a adaptabilidade de um único fluxo HSTCP para diversas

taxas de eventos de congestionamento. Quanto menor for a taxa de perdas de pacote, maior será a imparcialidade relativa.

Um importante aspecto a observar é que ambos os esquemas não estão prejudicando outros fluxos TCP, quando as perdas sistêmicas forem dominantes, porque ainda haverá banda disponível.

Quando perdas de pacote por congestionamento começam a emergir, a capacidade do enlace é alcançada e uma nova dinâmica para a imparcialidade aparece. Após este ponto, a imparcialidade será determinada pela política de gerenciamento de enfileiramento do roteador, pelo volume de tráfego, pelos parâmetros de cada esquema ("N" para os fluxos paralelos e "Low_Window, High_Window e High_P" para o HSTCP). Após este ponto de mudança, os fluxos paralelos começam a roubar banda dos outros fluxos TCP que competem pela banda. A quantidade roubada é proporcional ao número de fluxos paralelos empregados e ao volume do tráfego concorrente.

Quando um fluxo HSTCP é empregado, a quantidade de banda roubada é função dos seus parâmetros e também do volume do tráfego concorrente. A quantidade de banda roubada decresce quando o tráfego concorrente aumenta. A influência do gerenciamento de enfileiramento do roteador é expressado na Figura 5.19(a) e na Figura 5.19(b). Estes gráficos mostram que para usar a mesma quantidade de banda é requerido um número maior de fluxos REGTCP quando DT é empregado.

Deve ser notado também que um fluxo HSTCP perde mais de sua porção de banda que uma transmissão com fluxos paralelos, quando ocorre uma variação no tráfego concorrente. Isto está expresso na inclinação das linhas na Figura 5.28. Quando o tráfego concorrente muda, também o faz a taxa de eventos de congestionamento.

Este é um ponto importante a ser considerado quando uma transmissão volumosa de dados é executada em um enlace congestionado. A maior variação apresentada pelo HSTCP, comparada com a variação que ocorre com os fluxos paralelos, permite-o adaptar-se rapidamente a uma variação no tráfego e na taxa de eventos de congestionamento.

Como discutido anteriormente, não existe uma clara desvantagem em se usar o HSTCP quando comparado com fluxos paralelos para uma transferência volumosa de dados. Mesmo a limitação no tamanho máximo do *buffer* TCP pode ser um problema em

comum se um número pequeno de fluxos paralelos for usado. A possibilidade de mudar-se apenas a pilha TCP, ao invés de se alterar os programas já em uso, é muito atrativa. A imparcialidade é uma preocupação comum a ambos os esquemas, entretanto, nossa opinião é que o HSTCP apresenta uma melhor adaptabilidade a ambientes de taxa de eventos de congestionamento variáveis. Os fluxos paralelos também podem apresentar uma melhor adaptabilidade (usando o controle de congestionamento fracionário ou algum outro tipo de controle adaptativo), mas ao custo de perder sua simplicidade.

Apresentaremos nos parágrafos seguintes os experimentos desenvolvidos para observar os aspectos de desempenho e imparcialidade destes esquemas de transferência volumosa de dados.

Comparação por Simulação dos Esquemas de Transferência Volumosa de Dados

Dois experimentos foram desenvolvidos para comparar o emprego dos fluxos paralelos e do HSTCP. O primeiro trata com as reações a um enlace com perdas e o segundo com o comportamento quando um tráfego em rajadas está presente. Em ambos experimentos, 10 fluxos REGTCP estão presentes para medir o impacto de ambos os esquemas de transferência volumosa de dados nestes fluxos de longa duração.

FLUXOS PARALELOS EM CONDIÇÃO DE ENLACE COM PERDAS

O desempenho dos fluxos paralelos em um ambiente com perdas de pacote sistêmicas é definido pela taxa de eventos de congestionamento e pelo número de fluxos paralelos, conforme dito anteriormente, e confirmado na Figura 5.31(a). Após o limite de banda ter sido alcançado, a utilização do enlace permanece constante. O mesmo comportamento é observado para um fluxo HSTCP no mesmo gráfico. Para taxas de perda maiores que 10^{-3}, ambos os esquemas possuem um baixo desempenho. Este baixo desempenho é o resultado de uma grande aumento na taxa de eventos de congestionamento, principalmente devido à retransmissão de pacotes, conforme ilustrado na Figura 5.32(a).

O impacto do uso de fluxos paralelos e HSTCP em fluxos de longa duração é apresentado na Figura 5.30(a). Este gráfico mostra a utilização do enlace agregada de 10 fluxos REGTCP quando ambos os esquemas de transferência volumosa de dados

são empregados, e também quando não existe esta interferência. Não existe diferença no desempenho dos 10 fluxos com e sem competição antes da banda total do enlace ter sido atingida. Ambos os conjuntos de tráfego (10 fluxos REGTCP de longa duração e o tráfego concorrente de transferência volumosa de dados) tem espaço para crescer. Após a banda estar completamente utilizada, não existe variação na quantidade de banda usada pelos 10 fluxos REGTCP.

A Figura 5.33(a) mostra quantas vezes mais de banda, um conjunto de fluxos paralelos, está usando do que um único fluxo REGTCP (um décimo da utilização do enlace agregada de todos os 10 fluxos REGTCP). Está claro que esta razão é aproximadamente constante durante uma ampla faixa de taxas de perdas no enlace. Este comportamento apenas muda quando existe uma taxa de perdas de pacote muito pesada, superior a 10^{-4}. Em comparação, a mesma razão não é constante quando o HSTCP é usado. Ela muda com o nível de taxa de perdas no enlace, porque o HSTCP foi projetado para atuar desta maneira. Este comportamento variável reflete sua adaptabilidade a um ambiente com variação nas perdas no enlace. Ele pode atuar como um único fluxo REGTCP quando a taxa de perdas no enlace está em torno de 10^{-3} e como 5 fluxos REGTCP quando a taxa de perdas no enlace está em torno de 10^{-5}. Esta adaptabilidade representa uma vantagem em relação ao uso de fluxos paralelos, porque é difícil saber antecipadamente qual será a taxa de perda de pacote sistêmica mínima e a máxima banda disponível, de forma a não prejudicar muito os outros fluxos TCP que competem pela banda.

Quando o gerenciamento de enfileiramento DT é usado, a diferença de desempenho em relação ao RED apenas acontece após o limite da banda ter sido atingido. Também, a utilização do enlace é mantida similar para os fluxos paralelos após o limite de banda ter sido atingido, Figura 5.30(b).

DT permite uma maior agressividade por parte do HSTCP. O HSTCP toma mais banda dos 10 fluxos REGTCP, como visto na Figura 5.30(b), e conseqüentemente apresenta uma ampla faixa de imparcialidade relativa, como mostrado na Figura 5.33(b).

FLUXOS PARALELOS NA CONDIÇÃO DE TRÁFEGO EM RAJADAS

O desempenho dos métodos de transferência volumosa de dados é pressionado quando eles rodam em um ambiente com tráfego em rajadas. O desempenho dos fluxos paralelos tende a decrescer quando o número de perturbações aumenta, como visto na

Figura 5.35(a). Este comportamento é o mesmo apresentado por outros fluxos usando REGTCP quando submetidos a situações similares, como mostrado na Figura 5.13(a).

O desempenho de um fluxo HSTCP é muito menos sensível a este ambiente. De fato, o desempenho do HSTCP, quando competindo contra 10 fluxos REGTCP, melhora quando o número de perturbações aumenta. Isto é claro na Figura 5.35(a) e Figura 5.36(a). A primeira conclusão poderia ser que o HSTCP está roubando banda dos fluxos REGTCP, mas a Figura 5.37(a) mostra que a quantidade roubada decresce com o aumento do número de fluxos em rajada. A única explicação possível para este comportamento é que o HSTCP está usando a porção de banda deixada pelos fluxos REGTCP, devido eles estarem sendo prejudicados pelo tráfego em rajada. Isto representa uma excelente característica para o HSTCP neste tipo de ambiente (tráfego em rajadas e gerenciamento de enfileiramento do roteador RED), quando comparado os fluxos paralelos: o HSTCP é capaz de usar a porção de banda deixada pelos fluxos REGTCP quando submetidos a tráfego em rajadas. Fluxos paralelos não são capazes disto.

Para o método de fluxos paralelos, a imparcialidade relativa é mantida quase constante ao longo desta faixa de número de perturbações e é proporcional ao número de fluxos TCP paralelos. Como dito acima, para o caso em que o HSTCP está presente, a imparcialidade relativa aumenta quando o número de perturbações aumenta, e o fluxo HSTCP é capaz de usar mais banda. Estes fatos estão apresentados na Figura 5.36(a). Como salientado anteriormente, fica claro que o HSTCP tem uma melhor adaptabilidade e usa mais a banda, sem prejudicar significativamente os outros fluxos.

Esta situação muda quando o DT é usado para o gerenciamento de enfileiramento do roteador. O desempenho reduzido dos fluxos paralelos para um grande número de perturbações não é claro, pelo menos para o número de perturbações usadas neste experimento, Figura 5.35(b). Portanto, o desempenho dos fluxos paralelos é mantido quase constante quando o número de perturbações aumenta. O comportamento do HSTCP segue o mesmo comportamento. A única diferença em relação ao RED é que o HSTCP usa muito mais banda.

A imparcialidade relativa permanece sem mudanças para os fluxos paralelos, conforme havia sido para o caso do RED. Para o HSTCP, a imparcialidade relativa pode se espalhar ao longo de uma ampla faixa de valores, como visto na Figura 5.36(b). Muito embora um certo grau de incerteza esteje presente na imparcialidade relativa, a utilização do enlace pelo HSTCP é relativamente constante, Figura 5.35(b), e a quantidade de

banda roubada dos fluxos REGTCP competindo com o fluxo HSTCP, decresce quando o número de perturbações é alto, conforme visto na Figura 5.37(b).

6.3 Outras Questões

Esta seção explora algumas outras questões observadas durante o desenvolvimento deste trabalho. Tais questões não estão diretamente relacionadas com o tema principal de investigação, mas tiveram um certo impacto nos resultados.

6.3.1 Problema da Partida Lenta

O algoritmo padrão da Partida Lenta fornece um crescimento exponencial no tamanho da janela de congestionamento, dobrando o seu tamanho cada vez que um pacote ACK é recebido. A razão para este rápido crescimento é realizar uma rápida sondagem da capacidade da rede e começar rapidamente a transmitir próximo da capacidade do enlace. Todavia, dobrar a janela de congestionamento a cada RTT, em um enlace de alta capacidade, pode facilmente resultar em milhares de pacotes sendo descartados em um único tempo de percurso. Este descarte de um grande número de pacotes pode resultar em *timeouts* de retransmissão desnecessários para uma conexão TCP. A conexão TCP pode começar a fase de Prevenção de Congestionamento com uma janela de congestionamento muito baixa, levando um grande número de RTTs para recuperar sua antiga janela de congestionamento [17]. Portanto, o algoritmo tradicional de Partida Lenta pode ocasionar um baixo desempenho para grandes janelas de congestionamento.

Este problema poderia afetar seriamente os resultados de desempenho do presente trabalho. Este comportamento acontece tanto com o HSTCP quanto com o REGTCP pois ambos usam o mesmo algoritmo de Partida Lenta. Muitas vezes isto afeta o desempenho dos fluxos em estado estacionário, impedindo que os fluxos utilizem uma banda elevada.

Este problema foi identificado durante a realização experimentos descritos neste trabalho e o *draft* Partida Lenta Limitada para TCP com Grandes Janelas de Congestionamento (*Limited Slow-Start for TCP with Large Congestion Windows*) [17] foi proposto

como uma solução. A Partida Lenta Limitada foi usada neste trabalho.

O parâmetro MAX_SSTHRESH foi introduzido e a Partida Lenta foi modificada para grandes janelas de congestionamento, como mencionado na sessão 2.3.

Foi preciso sintonizar o parâmetro MAX_SSTHRESH para ter um desempenho razoável na Partida Lenta e evitar ter fluxos quase paralisados no começo de sua transmissão. Um MAX_SSTHRESH de 100 pacotes, conforme sugerido em [17], foi usado neste estudo.

O efeito de uma perda massiva de dados pode ser visto na Figura 5.39. Este gráfico mostra o efeito das perdas nos números de seqüência de um fluxo HSTCP usando diferentes valores para este parâmetro e diferentes algoritmos para a Partida Lenta.

A Tabela 5.1 e a Figura 5.38 dão a comparação entre várias condições de Partida Lenta. O MAX_SSTHRESH = 0 representa o algoritmo de Partida Lenta padrão. Está claro que mesmo um fluxo com algoritmo padrão alcançando a fase de prevenção de congestionamento mais rapidamente que os demais, ele terá uma grande perda de pacotes.

6.3.2 Vizinhança do Limite de Banda

A Figura 3.1 mostra a função resposta teórica do HSTCP, quando submetido a diferentes taxas de eventos de congestionamento. Ela também apresenta a capacidade do enlace usado neste trabalho. A intersecção entre estas duas linhas define o desempenho teórico máximo de 1 fluxo HSTCP neste estudo.

Neste trabalho foi possível ver que um fluxo HSTCP pode estar próximo deste ponto. A partir dos dados coletados no experimento da Figura 5.1(a), foi possível dizer que um fluxo HSTCP teve 1 evento de congestionamento a cada 638.000 pacotes enviados, em aproximadamente 7,73 segundos. Isto representa uma taxa de eventos de congestionamento de 1,56*10^{-6} e uma taxa de transferência de 8.244 pacotes/RTT, perto da banda limite de 8.333 pacotes/RTT, ou 98,9% de utilização da banda.

6.3.3 Problema na Implementação do TCP SACK para Grandes Janelas de Congestionamento

O TCP de Alta Velocidade está projetado para ser executado num regime de grandes janelas de congestionamento. Pretende-se atingir esta condição em redes futuras. Por esta razão, a implementação do algoritmo do TCP dificilmente foi testada nesta condição e nenhuma experiência foi desenvolvida.

Durante o desenvolvimento deste trabalho, uma permanente perda de desempenho foi observada durante o emprego do HSTCP, particularmente quando uma perda de pacote ocorria em uma grande janela de congestionamento. O sintoma era um duplo corte na janela de congestionamento, o primeiro corte ocorre quando a janela de congestionamento está próxima a 25.000 pacotes e o próximo corte em torno de 20.000 pacotes, como visto na Figura 6.1.

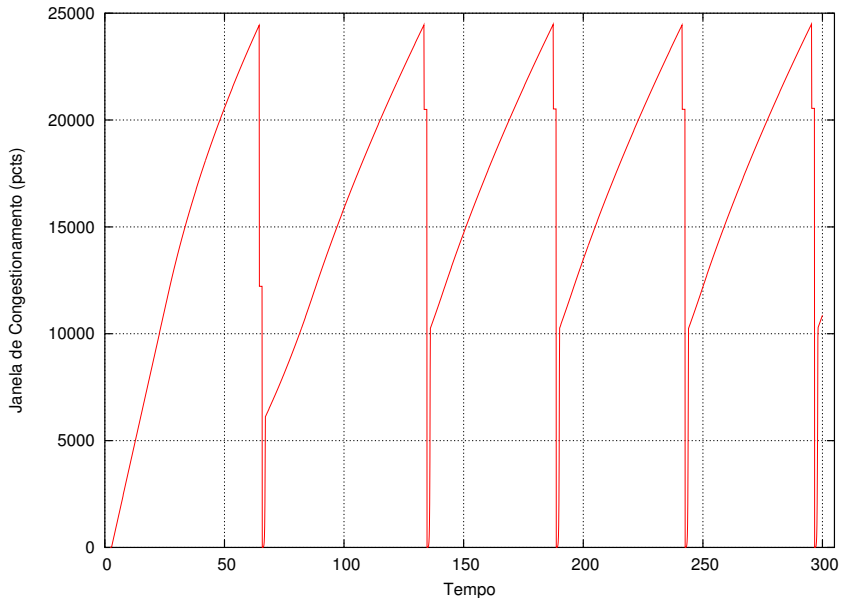


Figura 6.1: Evolução da Janela de Congestionamento - HSTCP Defeituoso - DT

Após longa investigação, notou-se que a implementação do TCP SACK no simulador NS-2, antes da versão ns2-1b9a (tcp.cc 1.52, tcp-sink.cc 1.46 e scoreboard.cc 1.14) não reportava corretamente o número de blocos SACK para janelas superiores a 1000 pacotes.

Embora este defeito tenha sido solucionado, uma solução escalável ainda não foi implementada. A atual implementação trabalha de acordo com a respectiva RFC, mas, para conexões TCP de alta velocidade onde CWND é superior a 10.000 pacotes, a rapidez da simulação reduz-se dramaticamente.

Este problema ressalta a crescente dificuldade de se ter uma implementação em operação para redes de alta velocidade. O algoritmo usado em implementações prévias pode não se adequar bem em cenários de alto desempenho. O segundo aspecto a observar é que certos problemas de implementação são apenas notados em cenários específicos e a complexidade da implementação do TCP é um desafio.

Capítulo 7

Conclusão e Trabalhos Futuros

A pressão por banda de rede mais larga tem produzido inovações tecnológicas que têm aumentado em várias vezes a capacidade de banda disponível. Novas tecnologias, tais como enlaces óticos têm possibilitado transferir diversos terabytes de informação em um tempo relativamente pequeno. Todavia, o TCP tem dificuldade em atingir plena utilização destes enlaces óticos, particularmente em conexões de longa distância. Portanto, várias aplicações de rede são incapazes de tirar vantagem destas novas redes de alta velocidade e utilizar plenamente a banda disponível.

O propósito deste trabalho foi estudar o emprego da proposta do TCP de Alta Velocidade em redes de alta velocidade e longa distância. Nós trabalhamos em torno de quatro pontos principais:

- o comportamento do TCP de Alta Velocidade em situações onde o TCP Padrão tem fraco desempenho;
- a possibilidade de usar o TCP de Alta Velocidade em conjunto com o TCP Padrão, mantendo a imparcialidade;
- o efeito da política de enfileiramento do roteador no desempenho do TCP de Alta Velocidade e na imparcialidade entre o TCP de Alta Velocidade e o TCP Padrão; e
- a possibilidade de usar o TCP de Alta Velocidade como um substituto para outras soluções existentes de transferência volumosa de dados.

Como resultado deste trabalho, nós concluímos que o TCP de Alta Velocidade de fato tem um desempenho melhor que o TCP Padrão para enlaces de alta velocidade e longa distância. O TCP de Alta Velocidade aumenta sua taxa de transferência mais rapidamente e sua recuperação de um evento de congestionamento leva menos tempo. Estas características aumentam a sua utilização de enlace.

Nós também mostramos que a porção de banda usada pelos fluxos do TCP de Alta Velocidade foi maior que a usada por fluxos do TCP Padrão, quando ambos os tipos de fluxo competiam pelo mesmo enlace. Entretanto, ficou nítido que a proporção de banda usada pelos fluxos do TCP de Alta Velocidade decrescia quando o total de eventos de congestionamento aumentava.

A troca do esquema de gerenciamento de enfileiramento do roteador não afetou significativamente a utilização do enlace pelo TCP de Alta Velocidade na maioria dos casos. O padrão geral de imparcialidade relativa encontrado quando RED foi usado, também foi seguido quando DT foi empregado. A diferença foi na maior quantidade de banda que os fluxos do TCP de Alta Velocidade retiraram dos fluxos do TCP Padrão, quando DT era usado.

Nós constatamos que o TCP de Alta Velocidade requer apenas a troca na pilha TCP usada nos transmissores. Isto representa uma vantagem sobre outros tipos de transferência volumosa de dados, como por fluxos paralelos, onde é necessária a mudança nos programas e o estabelecimento antecipado do número de fluxos paralelos a usar. Imparcialidade é uma preocupação no emprego do TCP de Alta Velocidade. Entretanto, é nossa opinião que o TCP de Alta Velocidade apresenta uma melhor adaptabilidade a ambientes com taxa de eventos de congestionamento variável.

Alguns passos podem seguir este trabalho. O mais importante agora é observar o comportamento do TCP de Alta Velocidade em redes de alta velocidade reais. Alguns testes já tiveram início [9].

Nós antevemos que um estudo sobre os parâmetros usados no TCP de Alta Velocidade seria necessário, para explorar diferentes combinações de valores. A troca nos parâmetros irá modificar a função resposta para diferentes níveis de taxa de eventos de congestionamento. Sua agressividade e imparcialidade, em comparação com o TCP Padrão serão afetadas. Outras funções, além da linear, podem ser testadas.

O comportamento em transientes é outra área de investigação. A reação do TCP de Alta Velocidade a súbita mudança de banda disponível é importante para averiguar seu tempo de recuperação e estabilização. Este exame foi feito indiretamente neste trabalho, mas existe espaço para mais investigação.

Uma questão a ser explorada é se a modificação dos algoritmos de controle de congestionamento do TCP para melhorar a taxa de transferência, como é feito no TCP de Alta Velocidade, é mais atraente do que simplesmente modificar o tamanho padrão do quadro Ethernet. Atualmente, o tamanho do quadro está congelado em 1500 bytes. Este tamanho de quadro representa a maioria dos tamanhos de pacote usados na Internet atualmente. O atual tamanho de quadro da Ethernet restringe a taxa de transferência em enlaces de alta velocidade e alto atraso.

Um ponto final a pensar é sobre uma revisão do conceito de imparcialidade como é usado atualmente para ser *TCP-Amigável*. Como é possível exigir que uma nova implementação seja completamente imparcial com a versões clássicas do TCP se é sabido que estas versões tem um fraco desempenho em enlaces com um alto produto banda atraso? Neste caso, manter a imparcialidade também significa manter um fraco desempenho. Evidentemente isto não é aceitável.

Os reais benefícios de aumentar a disponibilidade de enlaces de longa distância com capacidade de gigabits não será completamente realizada para aplicações que demandem alta banda, tais como transferência volumosa de dados, transmissão multimídia e grades computacionais, se o principal protocolo de transporte não for capaz de usar totalmente a capacidade de enlace disponível.

Sendo este um dos primeiros trabalhos a averiguar o comportamento do TCP de Alta Velocidade em diversos ambientes de rede, ele fornece uma base inicial de comparação para outras formas de avaliação que poderão ser usadas. Os dados numéricos obtidos por simulação poderão ser confrontados com levantamentos realizados em redes reais, possibilitando a confrontação das implementações do TCP de Alta Velocidade no simulador e na pilha TCP de uma máquina hospedeira. Da mesma forma, este estudo possibilita o aprimoramento de um modelo analítico do desempenho do TCP de Alta Velocidade, observando-se o comportamento do protocolo em diversas situações de rede. Desta maneira acreditamos que este trabalho contribuiu para o entendimento e avaliação do TCP de Alta Velocidade, e questões relacionadas ao desempenho do TCP em enlaces de longa distância.

Bibliografia

[1] M. Allman, V. Paxson, and R. Stevens. TCP congestion control. Internet Engineering Task Force, Abril 1999. RFC2581.

[2] J. Bolot. Characterizing end-to-end packet delay and loss in the internet. *Journal of High Speed Networks*, 2(3):289–298, Setembro 1993.

[3] T. Bonald, M. May, and J. Bolot. Analytic evaluation of RED performance. In *Proceedings of the 2000 IEEE Computer and Communications Societies Conference on Computer Communications*, pages 1415–1424, Agosto 2000.

[4] B. Braden, D. Clark, J. Crowcroft, B. Davie, S. Deering, D. Estrin, S. Floyd, V. Jacobson, G. Minshall, C. Partridge, L. Peterson, K. Ramakrishnan, S. Shenker, J. Wroclawski, and L. Zhang. Recommendations on queue management and congestion avoidance in the Internet. Internet Engineering Task Force, Abril 1998. RFC2309.

[5] R. Braden. Requirements for internet hosts - communication layers. Internet Engineering Task Force, Outubro 1989. RFC1122.

[6] L. Breslau, D. Estrin, K. Fall, S. Floyd, J. Heidemann, A. Helmy, P. Huang, S. McCanne, K. Varadhan, Y. Xu, and H. Yu. Advances in network simulation. *IEEE Computer*, 33(5):59–67, Maio 2000.

[7] R. Carlson. Tackling the end-to-end performance problem. Large Scale Networking Workshop, Março 2001. URL http://www.ngi-supernet.org/lsn2000/Argonne_Natl_Lab-Carlson.pdf.

[8] D. Chiu and R. Jain. Analysis of the increase and decrease algorithms for congestion avoidance in computer networks. *Journal of Computer Networks and ISDN Systems*, 17(1):1–14, Junho 1989.

[9] F. Coccetti and L. Cottrell. TCP stacks comparison with a single stream. Stanford Linear Accelerator Center, Março 2003. URL http://www-iepm.slac.stanford.edu/monitoring/bulk/fast/tcp-comparison.html.

[10] Euclidian Consulting. Description and resolution to reported packet loss problem with toshiba pcx1000 cable modems. URL http://www.cablemodemhelp.com/pcx1000.htm.

[11] T. Duningan, M. Mathis, and B. Tierney. A TCP tuning daemon. In *Procedings of IEEE Super Computing 2002*, 2002.

[12] P. Dykstra. Gigabit ethernet jumbo frames, Dezembro 1999. URL http://sd.wareonearth.com/˜phil/jumbo.html.

[13] K. Fall and S. Floyd. Simulation-based comparisons of Tahoe, Reno and SACK TCP. *Computer Communication Review*, 26(3):5–21, 1996.

[14] M. Fisk and W. Feng. Dynamic right-sizing in TCP. In *Procedings of Los Alamos Computer Science Institute Symposium*, Outubro 2001.

[15] S. Floyd. Congestion control principles. Internet Engineering Task Force, Setembro 2000. RFC2914.

[16] S. Floyd. Recommendation on using the gentle variant of RED, Março 2000. URL http://www.icir.org/floyd/red/gentle.html.

[17] S. Floyd. Limited slow-start for TCP with large congestion windows, Maio 2002. Internet draft draft-floyd-tcp-slowstart-00b.txt, work in progress.

[18] S. Floyd. Highspeed TCP for large congestion window, Fevereiro 2003. Internet Draft draft-floyd-tcp-highspeed-01.txt, work in progress.

[19] S. Floyd and K. Fall. Promoting the use of end-to-end congestion control in the Internet. *IEEE/ACM Transactions on Networking*, 7(4):458–472, Agosto 1999.

[20] S. Floyd, R. Gummadi, and S. Shenker. Adaptive RED: An algorithm for increasing the robustness of RED's active queue management. URL http://www.icir.org/floyd/papers/adaptativeRed.pdf, Agosto 2001.

[21] S. Floyd and V. Jacobson. On traffic phase effects in packet-switched gateways. *Internetworking: Research and Experience*, 3(3):115–156, Setembro 1992.

[22] S. Floyd and V. Jacobson. Random early detection gateways for congestion avoidance. *IEEE/ACM Transactions on Networking*, 1(4):397–413, Agosto 1993.

[23] S. Floyd, S. Ratnasamy, and S. Shenker. Modifying TCP's congestion control for high speeds. Preliminary Draft. URL http://www.icir.org/floyd/papers/hstcp.pdf, 2002.

[24] S. Gadde, J. S. Chase, and A. M. Vahdat. Coarse-grained network simulation for wide-area distributed systems. In *Communication Networks and Distributed Systems Modeling and Simulation Conference*, Janeiro 2002.

[25] T. Hacker, B. Athey, and B. Noble. The end-to-end performance effects of parallel TCP sockets on a lossy wide-area network. In *16th International Parallel and Distributed Processing Symposium*, Abril 2002. URL http://www.cnaf.infn.it/ferrari/papers/tcp/IPDPS.pdf.

[26] T. J. Hacker, B. D. Noble, and B. D. Athey. The effects of systemic packet loss on aggregate TCP flows. In *IEEE/ACM Supercomputing 2002: High Performance Networking and Computing*, Novembro 2002.

[27] G. Hasegawa and M. Murata. Survey on fairness issues in TCP congestion control mechanisms. *IEICE Transactions on Communications*, E84-B(6):1461–1472, Junho 2001.

[28] W. Huntoon, T. Dunigan, and B. Tierney. The Net100 project: Development of network-aware operating systems. URL http://www.net100.org.

[29] Linux Headquarters Inc. Linux kernel 2.4 variables, 2002. URL http://www.linuxhq.com/kernel/v2.4/doc/networking/ip-sysctl.txt.html.

[30] B. Irwin and M. Mathis. Web100: Facilitating high-performance network use, Janeiro 2001. URL http://www.internet2.edu/E2E/papers/20010109-E2EPM-Irwin.pdf.

[31] V. Jacobson. Congestion avoidance and control. In *Proceedings of the ACM SIGCOMM '88 Conference on Communications Architectures and Protocols*, volume 18, pages 314–329, Stanford, CA, Agosto 1988.

[32] V. Jacobson, R. Braden, and D. Borman. TCP extensions for high performance. Internet Engineering Task Force, Maio 1992. RFC1323.

[33] C. Jin, D. Wei, S. H. Low, G. Buhrmaster, J. Bunn, D. H. Choe, R. L. A. Cottrell, J. C. Doyle, W. Feng, O. Martin, H. Newman, F. Paganini, S. Ravot, and S. Singh. FAST TCP: From theory to experiments. Submitted to IEEE Comunications Magazine, Internet Technology Series, Abril 2003.

[34] W. Johnston, W. Kramer, J. Leighton, and C. Catlett. A vision for DOE scientific networking driven by high impact science, Março 2002. URL http://www.lbl.gov/CS/html/Network_Vision_Whitepaper.pdf.

[35] D. Katabi, M. Handley, and C. Rohrs. Internet congestion control for high bandwidth-delay product networks. In *ACM Sigcomm 2002*, Agosto 2002.

[36] J. Kulik, R. Coulter, D. Rockwell, and C. Partridge. Paced TCP for high delay-bandwidth networks. In *Proceedings of IEEE Globecom*, Dezembro 1999.

[37] K. Kurata, G. Hasegawa, and M. Murata. Fairness comparisons between TCP Reno and TCP Vegas for future deployment of TCP Vegas. In *Proceedings of IEEE INET 2000*, Julho 2000.

[38] T. Lakshman and U. Madhow. Performance analysis of window-based flow control using TCP/IP: Effect of high bandwidth-delay products and random loss. In *Fifth International Conference on High Performance Networking*, pages 135–149, Junho 1994.

[39] J. Lee, D. Gunter, B. Tierney, W. Allock, J. Bester, J. Bresnahan, and S. Tuecke. Applied techniques for high bandwidth data transfers across wide area network. In *Computing in High Energy and Nuclear Physics*, Beijing, China, Abril 2001. LBNL-46269.

[40] W. Leland, M. Taqqu, W. Willinger, and D. Wilson. On the self-similar nature of ethernet traffic (extended version). In *IEEE/ACM Transactions on Networking*, volume 2, pages 1–15, 1994.

[41] S. Low, L. Peterson, and L. Wan. Understanding TCP Vegas: A duality model. In *Proceedings of ACM SIGMETRICS 2001*, Cambridge, USA, Junho 2001.

[42] M. Mathis, J. Mahdavi, S. Floyd, and A. Romanow. TCP selective acknowledgment options. Internet Engineering Task Force, Outubro 1996. RFC2018.

[43] M. Mathis, J. Semke, J. Mahdavi, and T Ott. The macroscopic behavior of the TCP congestion avoidance algorithm. *Computer Communications Review*, 27(3), Julho 1997.

[44] S. McCreary and K. Claffy. Trends in wide area ip traffic patterns - a view from ames internet exchange. In *13th ITC Specialist Seminar on Internet Traffic Measurement and Modelling*, Monterey, CA, Setembro 2000.

[45] J. Nagle. Congestion control in IP/TCP Internetworks. Internet Engineering Task Force, Janeiro 1984. RFC896.

[46] R. Nitzan and B. Tierney. Experiences with TCP/IP over an ATM OC12 WAN. Technical Report LBNL-44765, Lawrence Berkeley National Laboratory, Abril 1999.

[47] C. Partridge. Buffer size = bw-delay product, part 2. End-to-End Mail List Archive, Abril 1998. URL ftp://ftp.isi.edu/end2end/end2end-interest-1998.mail.

[48] V. Paxson. *Measurements and Analysis of End-to-end Internet Dynamics*. PhD thesis, University of California at Berkeley, Abril 1997.

[49] K. Pentikousis, H. Badr, and B. Kharma. On the performance gains of TCP with ECN. In *Proceedings of the 2nd European Conference on Universal Multiservice Networks*, Colmar, France, Abril 2002.

[50] The VINT Project. NS-2 network simulator. URL http://www.isi.edu/nsnam/ns.

[51] K. Ramakrishnan and S. Floyd. A proposal to add explicit congestion notification (ECN) to IP. Internet Engineering Task Force, Janeiro 1999. RFC2481.

[52] K. Ramakrishnan, S. Floyd, and D. Black. The addition of explicit congestion notification (ECN) to IP. Internet Engineering Task Force, Setembro 2001. RFC3168.

[53] J. Salim and U. Ahmed. Performance evaluation of explicit congestion notification (ECN) in IP networks. Internet Engineering Task Force, Julho 2000. RFC2884.

[54] J. Semke, J. Mahdavi, and M. Mathis. Automatic TCP buffer tuning. *Computer Communication Review*, 28(4):315–323, Outubro 1998.

[55] M. Sooriyabandara and G. Fairhurst. Performance limitations due to TCP burstiness in GEO satellite networks with limited buffering. In *London Communications Symposium*, pages 127–130, London, UK, Setembro 2000.

[56] R. Stevens. *TCP/IP Illustrated, Volume 1: The Protocols*. Addison-Wesley, 1994.

[57] J. Stone and C. Partridge. When the CRC and TCP checksum disagree. In *ACM SIGCOMM Computer Communication Review: Proc. of the conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, pages 309–319, Stockholm, Sweden, Agosto 2000.

[58] B. Tierney. TCP tuning guide for distributed aplications on wide area networks. Technical Report LBNL-45261, Lawrence Berbeley National Laboratory, Fevereiro 2001.

[59] B. Welch. *Practical Programming in Tcl and Tk*. Prentice-Hall Inc., Upper Sadle River, New Jersey, USA, 2000.

[60] B. White, J. Lepreau, L. Stoller, R. Ricci, S. G. M. Newbold, M. Hibler, C. Barb, and A. Joglekar. An integrated experimental environment for distributed systems and networks. In *Fifth Symposium on Operating System Design and Implementation*, 2002.

[61] J. Winder. Equation-based congestion control. Master's thesis, University of Manheim, Fevereiro 2000.

[62] Y. Yang and S. Lam. General AIMD congestion control. Technical Report TR-200009, Department of Computer Science, University of Texas at Austin, Maio 2000. URL http://www.cs.utexas.edu/users/lam/NRL/TechReports.

[63] L. Zhang, S. Shenker, and D. Clark. Observations on the dynamics of a congestion control algorithm: The effects of two way traffic. In *Proceedings of ACM SIGCOMM'91*, pages 133–148, Zurich, Switzerland, Setembro 1991.